

STIC-Biotech/ChemLib

From: Zeman, Robert
Sent: Thursday, February 15, 2001 11:22 AM
To: STIC-Biotech/ChemLib
Subject: Search for 09/461,580

Please search SEQ ID NO: 1, 4, 9, 12, 19 and 25-26. Please include an interference search.

Thanks

Robert A. Zeman
Patent Examiner, USPTO Art Unit 1645
1911 South Clark Place, Rm 8D03
Arlington, VA 22202
(703) 308-7991

8E12
m3

CRFE

Point of Contact:
Toby Port
Technical Info. Specialist
CMI IE01 TEL: 308-3534

CWJ 1E01 1EG: 308-3234
Technical info. Specialist
Toru Hon
Point of Contact:

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Searcher: _____
 Searcher Phone #: _____
 Searcher Location: _____
 Date Searcher Picked Up: 2/16
 Date Completed: 2/27
 Searcher Prep & Review Time: 20
 Clerical Prep Time: _____
 Online Time: 10

Type of Search

NA Sequence (#) 1
 AA Sequence (#) 6
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems CS
 WWW/Internet _____
 Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:41:39 ; Search time 220.15 Seconds

(without alignments)
509.376 Million cell updates/sec

Title: US-09-461-580A-1

Perfect score: 3854

Sequence: 1 MADEVALALQAAGSPSAASAAAA.....NEAIAFRELITVDVNPSPDKS 737

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 936741 seqs, 152156132 residues

936741

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA:*

1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080.COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097.COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098.COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099.COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100.COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101.COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US102.COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US103.COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US104.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3854	100.0	737	US-09-461-580A-1	Sequence 1, Appl
2	3854	100.0	737	US-09-461-580A-26	Sequence 26, Appl
3	1732	44.9	327	US-09-461-580A-12	Sequence 12, Appl
4	1288	33.4	245	US-09-461-580A-4	Sequence 4, Appl
5	1169	30.3	823	US-09-619-049-576	Sequence 576, App

6	1169	30.3	823	22	US-60-167-217-10885	Sequence 10885, A
7	1169	30.3 <td>823</td> <td>22<th>US-60-171-627-915</th><th>Sequence 915, App</th></td>	823	22 <th>US-60-171-627-915</th> <th>Sequence 915, App</th>	US-60-171-627-915	Sequence 915, App
8	1169	30.3 <td>823</td> <td>22<th>US-60-173-464-8827</th><th>Sequence 8827, Ap</th></td>	823	22 <th>US-60-173-464-8827</th> <th>Sequence 8827, Ap</th>	US-60-173-464-8827	Sequence 8827, Ap
9	1169	30.3 <td>823</td> <td>22<th>US-60-191-637-10901</th><th>Sequence 10901, A</th></td>	823	22 <th>US-60-191-637-10901</th> <th>Sequence 10901, A</th>	US-60-191-637-10901	Sequence 10901, A
10	1169	30.3 <td>823</td> <td>22<th>US-60-191-681-8549</th><th>Sequence 8549, Ap</th></td>	823	22 <th>US-60-191-681-8549</th> <th>Sequence 8549, Ap</th>	US-60-191-681-8549	Sequence 8549, Ap
11	1169	30.3 <td>823</td> <td>22<th>US-09-270-767-44940</th><th>Sequence 44940, A</th></td>	823	22 <th>US-09-270-767-44940</th> <th>Sequence 44940, A</th>	US-09-270-767-44940	Sequence 44940, A
12	682	17.7 <td>128</td> <td>16<th>US-09-461-580A-9</th><th>Sequence 9, Appl1</th></td>	128	16 <th>US-09-461-580A-9</th> <th>Sequence 9, Appl1</th>	US-09-461-580A-9	Sequence 9, Appl1
13	643.5	16.7 <td>339</td> <td>16<th>US-09-248-796-19599</th><th>Sequence 19599, A</th></td>	339	16 <th>US-09-248-796-19599</th> <th>Sequence 19599, A</th>	US-09-248-796-19599	Sequence 19599, A
14	643.5	16.7 <td>339</td> <td>22<th>US-60-096-409-19599</th><th>Sequence 19599, A</th></td>	339	22 <th>US-60-096-409-19599</th> <th>Sequence 19599, A</th>	US-60-096-409-19599	Sequence 19599, A
15	568	14.7 <td>313</td> <td>16<th>US-09-248-796-20918</th><th>Sequence 20918, A</th></td>	313	16 <th>US-09-248-796-20918</th> <th>Sequence 20918, A</th>	US-09-248-796-20918	Sequence 20918, A
16	568	14.7 <td>313</td> <td>22<th>US-60-096-409-20918</th><th>Sequence 20918, A</th></td>	313	22 <th>US-60-096-409-20918</th> <th>Sequence 20918, A</th>	US-60-096-409-20918	Sequence 20918, A
17	567	14.7 <td>336</td> <td>18<th>US-09-461-580A-11</th><th>Sequence 11, Appl</th></td>	336	18 <th>US-09-461-580A-11</th> <th>Sequence 11, Appl</th>	US-09-461-580A-11	Sequence 11, Appl
18	556	14.4 <td>106</td> <td>18<th>US-09-461-580A-19</th><th>Sequence 19, Appl</th></td>	106	18 <th>US-09-461-580A-19</th> <th>Sequence 19, Appl</th>	US-09-461-580A-19	Sequence 19, Appl
19	549.5	14.3 <td>389</td> <td>22<th>US-60-169-629-517</th><th>Sequence 517, App</th></td>	389	22 <th>US-60-169-629-517</th> <th>Sequence 517, App</th>	US-60-169-629-517	Sequence 517, App
20	549.5	14.3 <td>389</td> <td>22<th>US-60-187-470-517</th><th>Sequence 517, App</th></td>	389	22 <th>US-60-187-470-517</th> <th>Sequence 517, App</th>	US-60-187-470-517	Sequence 517, App
21	535.5	13.9 <td>267</td> <td>18<th>US-09-461-580A-3</th><th>Sequence 3, Appl1</th></td>	267	18 <th>US-09-461-580A-3</th> <th>Sequence 3, Appl1</th>	US-09-461-580A-3	Sequence 3, Appl1
22	520	13.5 <td>355</td> <td>22<th>US-60-167-217-10557</th><th>Sequence 10557, A</th></td>	355	22 <th>US-60-167-217-10557</th> <th>Sequence 10557, A</th>	US-60-167-217-10557	Sequence 10557, A
23	520	13.5 <td>355</td> <td>22<th>US-60-173-464-8533</th><th>Sequence 8533, Ap</th></td>	355	22 <th>US-60-173-464-8533</th> <th>Sequence 8533, Ap</th>	US-60-173-464-8533	Sequence 8533, Ap
24	517	13.4 <td>272</td> <td>18<th>US-09-461-580A-2</th><th>Sequence 2, Appl1</th></td>	272	18 <th>US-09-461-580A-2</th> <th>Sequence 2, Appl1</th>	US-09-461-580A-2	Sequence 2, Appl1
25	491.5	12.8 <td>284</td> <td>16<th>US-09-248-796-19606</th><th>Sequence 19606, A</th></td>	284	16 <th>US-09-248-796-19606</th> <th>Sequence 19606, A</th>	US-09-248-796-19606	Sequence 19606, A
26	491.5	12.8 <td>284</td> <td>22<th>US-60-096-409-19606</th><th>Sequence 19606, A</th></td>	284	22 <th>US-60-096-409-19606</th> <th>Sequence 19606, A</th>	US-60-096-409-19606	Sequence 19606, A
27	484	12.6 <td>284</td> <td>22<th>US-60-117-905-57</th><th>Sequence 57, Appl</th></td>	284	22 <th>US-60-117-905-57</th> <th>Sequence 57, Appl</th>	US-60-117-905-57	Sequence 57, Appl
28	479	12.4 <td>257</td> <td>22<th>US-60-095-827-17</th><th>Sequence 17, Appl</th></td>	257	22 <th>US-60-095-827-17</th> <th>Sequence 17, Appl</th>	US-60-095-827-17	Sequence 17, Appl
29	454.5	11.8 <td>358</td> <td>18<th>US-09-417-507-39938</th><th>Sequence 39938, A</th></td>	358	18 <th>US-09-417-507-39938</th> <th>Sequence 39938, A</th>	US-09-417-507-39938	Sequence 39938, A
30	441.5	11.5 <td>230</td> <td>1<th>PCT-US00-02237-9</th><th>Sequence 9, Appl1</th></td>	230	1 <th>PCT-US00-02237-9</th> <th>Sequence 9, Appl1</th>	PCT-US00-02237-9	Sequence 9, Appl1
31	441.5	11.5 <td>230</td> <td>22<th>US-60-117-905-9</th><th>Sequence 9, Appl1</th></td>	230	22 <th>US-60-117-905-9</th> <th>Sequence 9, Appl1</th>	US-60-117-905-9	Sequence 9, Appl1
32	406	10.5 <td>183</td> <td>18<th>US-09-417-507-28899</th><th>Sequence 28899, A</th></td>	183	18 <th>US-09-417-507-28899</th> <th>Sequence 28899, A</th>	US-09-417-507-28899	Sequence 28899, A
33	367.5	9.5 <td>232</td> <td>18<th>US-09-461-580A-35</th><th>Sequence 35, Appl</th></td>	232	18 <th>US-09-461-580A-35</th> <th>Sequence 35, Appl</th>	US-09-461-580A-35	Sequence 35, Appl
34	328.5	8.5 <td>488</td> <td>16<th>US-09-248-796-19601</th><th>Sequence 19601, A</th></td>	488	16 <th>US-09-248-796-19601</th> <th>Sequence 19601, A</th>	US-09-248-796-19601	Sequence 19601, A
35	328.5	8.5 <td>488</td> <td>22<th>US-60-096-409-19601</th><th>Sequence 19601, A</th></td>	488	22 <th>US-60-096-409-19601</th> <th>Sequence 19601, A</th>	US-60-096-409-19601	Sequence 19601, A
36	312.5	8.1	107	18 <th>US-09-461-580A-20</th> <th>Sequence 20, Appl</th>	US-09-461-580A-20	Sequence 20, Appl
37	304	7.9	128	18 <th>US-09-461-580A-10</th> <th>Sequence 10, Appl</th>	US-09-461-580A-10	Sequence 10, Appl
38	295	7.7	245	18 <th>US-60-117-905-70</th> <th>Sequence 70, Appl</th>	US-60-117-905-70	Sequence 70, Appl
39	279.5	7.3	248	18 <th>US-09-450-969-6406</th> <th>Sequence 6406, Ap</th>	US-09-450-969-6406	Sequence 6406, Ap
40	278.5	7.2	270	16 <th>US-09-252-991A-19648</th> <th>Sequence 19648, A</th>	US-09-252-991A-19648	Sequence 19648, A
41	267.5	6.9	223	1 <th>PCT-US99-29950-188</th> <th>Sequence 188, App</th>	PCT-US99-29950-188	Sequence 188, App
42	267.5	6.9	223	19 <th>US-09-591-316-188</th> <th>Sequence 189, App</th>	US-09-591-316-188	Sequence 189, App
43	266.5	6.9	107	18 <th>US-09-461-580A-16</th> <th>Sequence 16, Appl</th>	US-09-461-580A-16	Sequence 16, Appl
44	265	6.9	253	20 <th>US-09-611-529-6441</th> <th>Sequence 6441, Ap</th>	US-09-611-529-6441	Sequence 6441, Ap
45	262	6.8	310	1 <th>PCT-US00-02237-25</th> <th>Sequence 25, Appl</th>	PCT-US00-02237-25	Sequence 25, Appl

ALIGNMENTS

US-09-461-580A-1

RESULT 1

: Sequence 1, Application US/09461580A

: GENERAL INFORMATION:

: APPLICANT: Guarente, Leonard

: APPLICANT: Imai, Shin-Ichiro

: APPLICANT: Armstrong, Christopher

: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH

: TITLE OF INVENTION: ALTERS HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE

: TITLE OF INVENTION: LIPSPAN

: FILE REFERENCE: 0050.1618-000

: CURRENT APPLICATION NUMBER: US/09/461.580A

: CURRENT FILING DATE: 1999-12-15

: NUMBER OF SEQ ID NOS: 35

: SOFTWARE: FastSeq for Windows Version 4.0

: SEQ ID NO 1

: LENGTH: 737

: TYPE: PRT

: ORGANISM: Mus musculus

US-09-461-580A-1

Query Match 100.0%; Score 3854; DB 18; Length 737;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 737; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MADEVALALQAAGSPSAASAAASOPADEPLRRDRGGLGRSPGPSAAVAAPAAAG 60

|||||

```

Db 1 MADEVALALQAAAGSPSAAAMEAASQPADEPLRKRRRDRDGPGLGSPGSPSAVAAPAAG 60
Oy 61 CEASAAAPAAALMREAAGAAASAREAPATAVAGDGNCSGLRREPRADDDEDEGE 120
Db 61 CEASAAAPAAALMREAAGAAASAREAPATAVAGDGNCSGLRREPRADDDEDEGE 120
Oy 121 DEAAAAAAGIYRDNLITDGLLTNGFHSCESDDDRTSHASSDWTPRRIGPYTFV 180
Db 121 DEAAAAAAGIYRDNLITDGLLTNGFHSCESDDDRTSHASSDWTPRRIGPYTFV 180
Oy 181 OOHLMIGDPTITLKDLPETIPPELDMTLMQIYINILSEPPKRRKKDINTIEDAVK 240
Db 181 OOHLMIGDPTITLKDLPETIPPELDMTLMQIYINILSEPPKRRKKDINTIEDAVK 240
Oy 241 LLOECKKIITVLGAGSVSCGIPDRSRDGIYARLAVDPDLDPQAMDIYFRKDRP 300
Db 241 LLOECKKIITVLGAGSVSCGIPDRSRDGIYARLAVDPDLDPQAMDIYFRKDRP 300
Oy 301 FFKFAKEIYPGQFOPSLCHKFTALSDKEGKLLRNTQNTIDLEQVAGIQRILOCHGSEAT 360
Db 301 FFKFAKEIYPGQFOPSLCHKFTALSDKEGKLLRNTQNTIDLEQVAGIQRILOCHGSEAT 360
Oy 361 ASCLCKKXVDEAVRGDIFNOVPRCPADPELAIMKPEIVFEGENLPEQFHRAMKY 420
Db 361 ASCLCKKXVDEAVRGDIFNOVPRCPADPELAIMKPEIVFEGENLPEQFHRAMKY 420
Oy 421 DKDEVDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLHLHFDVLLDCDVIINE 480
Db 421 DKDEVDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLHLHFDVLLDCDVIINE 480
Oy 481 LCHRLGGEYAKLCNPNVKLSEITEKPPRQKELVHLSLPPLHLHSEDSSPERTVPOD 540
Db 481 LCHRLGGEYAKLCNPNVKLSEITEKPPRQKELVHLSLPPLHLHSEDSSPERTVPOD 540
Oy 541 SSVIATLVDOATNNVNDLEVESSCVEKPOEVOISRVENINENPDKAVGSTADK 600
Db 541 SSVIATLVDOATNNVNDLEVESSCVEKPOEVOISRVENINENPDKAVGSTADK 600
Oy 601 NERTSVAETVRKCMRNRLAKEQISKRLBGNQYLFVPPNRYIFHGAEVYSDSDVLS 660
Db 601 NERTSVAETVRKCMRNRLAKEQISKRLBGNQYLFVPPNRYIFHGAEVYSDSDVLS 660
Oy 661 CGSNSDGTGQSPSLEPLEDESEIEEFYNGLEDTERECAGSGGFGADGGDOEVN 720
Db 661 CGSNSDGTGQSPSLEPLEDESEIEEFYNGLEDTERECAGSGGFGADGGDOEVN 720
Oy 721 IATROELTDVNPSPDKS 737
Db 721 IATROELTDVNPSPDKS 737

```

```

RESULT 2
US-09-461-580A-26
; Sequence 26, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-26

```

```

Query Match 100.0%; Score 3854; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 737; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MADEVALALQAAAGSPSAAAMEAASQPADEPLRKRRRDRDGPGLGSPGSPSAVAAPAAG 60
Db 1 MADEVALALQAAAGSPSAAAMEAASQPADEPLRKRRRDRDGPGLGSPGSPSAVAAPAAG 60
Oy 61 CEASAAAPAAALMREAAGAAASAREAPATAVAGDGNCSGLRREPRADDDEDEGE 120
Db 61 CEASAAAPAAALMREAAGAAASAREAPATAVAGDGNCSGLRREPRADDDEDEGE 120
Oy 121 DEAAAAAAGIYRDNLITDGLLTNGFHSCESDDDRTSHASSDWTPRRIGPYTFV 180
Db 121 DEAAAAAAGIYRDNLITDGLLTNGFHSCESDDDRTSHASSDWTPRRIGPYTFV 180
Oy 181 OOHLMIGDPTITLKDLPETIPPELDMTLMQIYINILSEPPKRRKKDINTIEDAVK 240
Db 181 OOHLMIGDPTITLKDLPETIPPELDMTLMQIYINILSEPPKRRKKDINTIEDAVK 240
Oy 241 LLOECKKIITVLGAGSVSCGIPDRSRDGIYARLAVDPDLDPQAMDIYFRKDRP 300
Db 241 LLOECKKIITVLGAGSVSCGIPDRSRDGIYARLAVDPDLDPQAMDIYFRKDRP 300
Oy 301 FFKFAKEIYPGQFOPSLCHKFTALSDKEGKLLRNTQNTIDLEQVAGIQRILOCHGSEAT 360
Db 301 FFKFAKEIYPGQFOPSLCHKFTALSDKEGKLLRNTQNTIDLEQVAGIQRILOCHGSEAT 360
Oy 361 ASCLCKKXVDEAVRGDIFNOVPRCPADPELAIMKPEIVFEGENLPEQFHRAMKY 420
Db 361 ASCLCKKXVDEAVRGDIFNOVPRCPADPELAIMKPEIVFEGENLPEQFHRAMKY 420
Oy 421 DKDEVDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLHLHFDVLLDCDVIINE 480
Db 421 DKDEVDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLHLHFDVLLDCDVIINE 480
Oy 481 LCHRLGGEYAKLCNPNVKLSEITEKPPRQKELVHLSLPPLHLHSEDSSPERTVPOD 540
Db 481 LCHRLGGEYAKLCNPNVKLSEITEKPPRQKELVHLSLPPLHLHSEDSSPERTVPOD 540
Oy 541 SSVIATLVDOATNNVNDLEVESSCVEKPOEVOISRVENINENPDKAVGSTADK 600
Db 541 SSVIATLVDOATNNVNDLEVESSCVEKPOEVOISRVENINENPDKAVGSTADK 600
Oy 601 NERTSVAETVRKCMRNRLAKEQISKRLBGNQYLFVPPNRYIFHGAEVYSDSDVLS 660
Db 601 NERTSVAETVRKCMRNRLAKEQISKRLBGNQYLFVPPNRYIFHGAEVYSDSDVLS 660
Oy 661 CGSNSDGTGQSPSLEPLEDESEIEEFYNGLEDTERECAGSGGFGADGGDOEVN 720
Db 661 CGSNSDGTGQSPSLEPLEDESEIEEFYNGLEDTERECAGSGGFGADGGDOEVN 720
Oy 721 IATROELTDVNPSPDKS 737
Db 721 IATROELTDVNPSPDKS 737

```

```

RESULT 3
US-09-461-580A-12
; Sequence 12, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12

```



```

0Y 444 PSSIPHEPOLLINREPPLHLPHEVELLGGCDYINELCHRLGSE---YAKLCNPYKLS 500
Db 439 PSSIPATYAPOLLINNEOHLHLKFEVELLGGSDYIINOICHRLLSPNDNCMLCCDESVLT 498
0Y 501 ETEKPPROKELVLSLSELPPTPLHISEDSSPPR-----TYPQDSSVATIWDQA 551
Db 499 ESKELMP--PEHSNHLHLH---HLHLHRCSESEEROSOLDITDIOISINSADYILISAG 554
0Y 552 TNNNVNDELVESSG-----VE-----ERKO--EVOT 576
Db 555 TCSO-SGSEESTFSGCKRSTAAEAALERIKTDLVLENETTALSCDRLEGGQTYES 6133
0Y 577 SRNV-----ENINVENPDEKAVGSSYADKN--ERTSVAETVRKCPWNRILAKDEISKRE 628
Db 614 YRHLISIDSKSGIEOCNEAETPSYVPRPSNLYOETKTVAESLPPIPOORGRQTAERLQ 6733
0Y 629 -GNQVLFPNRYIFHGAEV-----YDSSEDDVLSS-----SCGSN---SDSGTC 670
Db 674 PGTFYSHTNNYSYVFGAQVEMWDYSDDDDEEERSHNRHSDLFGVNGHNYKDDDEDAC 7333
0Y 671 Q-----SPSLDE-----EPLEDES 683
Db 734 DLNAVPLSLPLPSTLEAHITVDYINGSNELPLPNS 768

```

```

RESULT 6
US-60-167-217-10885
; Sequence 10885, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 10885
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-10885

```

Query Match	30.3%;	Score 1169;	DB 22;	Length 823;
Best Local Similarity	38.8%;	Pred. No. 3.1e-87;		
Matches 293;	Conservative 100;	Mismatches 206;	Indels 156;	Gaps 25;

OY	57	AAAGCEAASAAAPALMREAA-----GAASAREAPATAVAGODNSGRLRER-----	107
Dd	42	ASTSTEAALAEAAATTTTTPATSLAGKANEIKTKLLAREDEITANLEHTTKNPTKS	101
OY	108	AADFDDEGEDEEEDAAAAAAAIGYRDNLLLTGLLTNGFHSCESDDDRTSHASS-	166
Dd	102	MGEDEDEDEEEDDEEE-----EDD-----EGITGTSNEDSSSCSSSV	145
OY	167	-DWTPRRIGRPTFVOOHLMTIGTDPTILKDLPE--TTTPRLDMTLMQVINITSE	222
Dd	146	EPDKLR-----WLGREFTGKVPKQVYASINPHATGTAGLTDSDVIMDYATHLNE	198
OY	223	PKRRKRKDINTIEDAVKLLDCEKKIIVLTGAGVSYSCGIPDFERSRDGIARLAVDPDL	282
Dd	199	PKRRKRLASVTPFDVLSLVKSKRIIVLTGAGVSYSCGIPDFESTNGIARLADHPDL	258
OY	283	PDPQAPFIEYFKRDPREFFEAKETIYPGOGPSLCKPHTLASPSKEKLLRNTONTDL	342
Dd	259	PDPQAPFINTFKRDPREFFEAKETIYPGOGPSLCKPHTLASPSKEKLLRNTONTDL	318
OY	343	EQVAGIQRILQCHGSFATASCLICKYAVDCAEAVBGIDFNOVAVPCRPCP-----	391
Dd	319	ERVAGIQRVICHGSFATASCLICKYAVDCAEAVBGIDFNOVAVPCRPCP-----	378

```
QY      332 - ADDEPL-----ALMKREYFEGENLPEOCHNRMAKDKDCEVDLLIYTSSSLKYVRPAALI 443
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      379 AVTEBELQOLVENGIMKDIEDIVFEGEPDEXETWAMADKDDCDLLIYGSSLKVPRVAHI 438
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      444 PSSIPHEVPOLLINNEPRLPHHPVELLDGDDVIYNELCHRLOGE---YAKLCOPNKUS 500
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      439 PSSIPATYPOLLINREQHHLHKFDVELLGDSDVITINOICRLSNDNDCMQLCCDESVLFT 498
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      501 ETEKKPPROKELYHLSLPRLPHISEDSSPER-----TYPDODSVIATIIVDOA 551
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      499 ESKELMP-REHSNNHLLH---HLHHRRCSSESERQSOLDPTDQSIKSNSSADYLILGSAG 554
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      552 TNNANNDEVESSC-----VE-----EKPO-EYQT 576
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      555 TCSD-SGFESTFSOGKRSTAEAAALERIKTDLIVLENNETALSCLDRIGLEBPOTTVES 613
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      577 SRNV-----ENJVENPNDFKAVSGSSTAADKN---ERTSVAEVVKCMNPRLAKEOISKRE 628
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      614 YRHLSIDBSKSGIEFOCDNEHTPSYVRSPNLVOGTKTIVAFSLPIPIPOQRKROTAAERLO 673
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      639 -GNQYLFVPPNRYIFHGAEV-----YSDBEDDYLISS-----SCGSN---SDSGTC 670
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      674 PGTFFSHVNNTSYVFGAQVFWMDNDYSDDDDEEBERSHNHRSHDLFGVGWGHNYRDDDEDAC 733
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      671 Q-----SPSLE-----EPLEDSS 683
        :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      734 DUNAVPLPLPLPSPLEAHITVDIVNGSNELPJNKS 768
```

```

RESULT 7
US-60-171-627-915
; Sequence 915 Application US/60171627
; GENERAL INFORMATION:
; APPLICANT: Yandell, Mark
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO000177
; CURRENT APPLICATION NUMBER: US/60/171, 627
; CURRENT FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-171-627-915

```

Query Match	30.3%;	Score 1169;	DB 22;	Length 823;
Best Local Similarity	38.8%;	Pred. No. 3.1e-87;		
Matches 293;	Conservative 100;	Mismatches 206;	Indels 156;	Gaps 25;

QY	57	AALGGEAASAAAPALMREAA-----GAASAREAPATAVAGDGDNSGLRRE-----	107
Db	42	ASTSTEAEAEAPATYTTTPEATSELACAKANEITKTLAREOEIOGNIHEHKKNPTKS	101
QY	108	AADPDDEGEDEEAEAAAAAAAAAIGYRDNLLTLDGLTGNFHCESDDDDRTSHASS-	166
Db	102	MGEDEDEEEDEEEDEEE-----EDD-----EGEITGTSNDEDESSNCSSSV	145
QY	167	--DWTPRRIGRIFFFVOOHLMTGDPRIILKDLPE--TTPPELDMTLMQIYINILSE	222
Db	146	EPDMRLR-----WLOEFTYGRVPROVIASIMPHFATGLAGTDSVIMDYLAHLNE	198
QY	223	PKRRKRDDINTIEDAVVLLDECKIIITLTAGSVSGIPDFSRDGIYAARLADPEPD	282
Db	199	PKRRKLASVMTPEDDVISLVKKSOKIITLTAGSVSGCIGDPDFSTNGIYARLADPEPD	258
QY	283	PDPAQMFIDEYFRKDPRFFFAKEIYPGQFQPSLCHKFTALSDKECKLLRNTQNTIDL	342
Db	259	PDPAQMFIDINFKRDPREFYFAEIEIVGEEQSPSCRHFIEMETPKKLLRNTQNTIDL	318

```

OY 343 EOVAGIOIILLOCHSFPATSLACKYIVDCVAGDIENOVVPRCP- 391
Db 319 ERVAGIOIVIECHSFSSTASTCKCRFCMADLADIAFORIPVCPQCPQPNKQSVADY 378
OY 392 --ADEPL-----AIMKPEIVFEGENLPEQFHRAMKXDKDEVDLLIYGSSLKVRPVALI 443
Db 379 AVTEBELNLQVLNGENIMKRDIVFPEGGLPDEXHYTMATDKVCDLLIYGSSLKVRPVANH 438
OY 444 PSSIDPHEVPOILLINREPLPHLHEDVELLDGCDVIINELCHRLGE--YAKLCNPKVKS 500
Db 439 PSSIPATVPOILLINREOHLHLKFEVVELLIGSDVIINQICRLSDNDNCMQLCCDESIVLT 498
OY 501 EITEKPPROKVELHLSLPTPLPHIEDSSPBR-----TYPQDSVATITLVDA 551
Db 499 ESKELMP--PESHNLHLH--HLHLHRHCSESESOQLDPTQSIKSNSSADIYILGSAG 554
OY 552 TNNVNDLEVSSESC-----VE-----EKPO--EVOT 576
Db 555 TCSD--SGEESSTFCGKRSTAEAAATERIKTDLVLENETLATSCDRGLEBGPOTTVES 613
OY 577 SRNV-----ENINVENPDFFKAVGSSSTADKN--ERTSVAETVRKCMENRIAKQISKRL 628
Db 614 YRLHSIDSSKSGIEOCDCNEATPSYVRPSNLVQETKTVAAPSLAPIPOORGRKQTAERLD 673
OY 629 -GNQYLFVPRRYRYFHGAEV-----YRDSIEDVLISS-----SCGSN--SPSGTC 670
Db 674 PGTFSHTNNYSYVFPGAQVFWMDYDDDEEBEERHNHRHSDLFGVNGHNYKDDDEDAC 733
OY 671 Q-----SPSLE-----EPLEDES 683
Db 734 DLNAVPLSPLLPRLAHLIVTDIYNGSNEPLPNS 768

RESULT 8
US-60-173-464-8827
: Sequence 8827, Application US/60173464
: GENERAL INFORMATION:
: APPLICANT: LI, Peter W.D.
: TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
: FILE REFERENCE: CL000173
: CURRENT APPLICATION NUMBER: US/60/173,464
: CURRENT FILING DATE: 1999-12-29
: NUMBER OF SEQ ID NOS: 30269
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8827
: LENGTH: 823
: TYPE: PRT
: ORGANISM: Drosophila
: US-60-173-464-8827

```

	Query Match	30.3%	Score 1169	DB 22	Length 823	
	Best Local Similarity	38.8%	Pred.No.3.1e-87			
	Matches	293	Conservative	100	Mismatches	206; Indels 156; Gaps
Qy	57	AAAGCEAASAPAAPALWREA----	GAAASAREAPATAVAGDGDGSGLRREPR-----	107		
Db	42	ASTSTAEAEATATTTPATSELAGKANGEIKTKTLAAREQEIGANLEHKTKNPYS	101			
Qy	108	AADDDDEGEFEDEAAAAAAATGYRONLLLTGDLTNGFHSCSDDDDRSHASS--	166			
Db	102	MGEDDEDDEEEEDDEE-----	EDD-----	EETGTGSNEDDSSNCSSSV	145	
Qy	167	--DWTPRRRIGTYTEVOQHLMGTDPRTLIKDLPE--	TIPPELDMKLQIWINILSE	222		
Db	146	EPDMKLR-----	WLQREFYTGVRPQVIASIMPHFATGLAGDDDSVLMOYLARLNMB	198		
Qy	223	PPKKRRKDINTIEDAVKILQECKRIIVLTGAGVSVCISLPFRSMDGIYARLANDFPDL	282			
Db	199	PKRRKLASVNFEFDVISLVKRSOKITVLTGAGVSVCISLPFRSRNNGIYARLANHDFPDL	258			

[illegible]

	Query Match	30.3%	Score 1169:	DB 22:	Length 823:
	Best Local Similarity	38.8%	Pred No.3.le-87:		
	Matches 293:	Conservative 100:	Mismatches 206:	Indels 156:	Gaps 25:
Oy	AAAGCAASAAPALMTREA-----GAASAERAPATAVADGDGNGSLRPREP----	107			
Dd	: :				
	42 ASTSTEAELEAFATYTTTTPPATSELAGANGSIGIKKTLAAREQELAKTKTNPKS	101			
Oy	108 AADFDDEDDGEEDDEAAAAAAAAAIGYRDNLLLTLDGLTFNGHSCSDODDRTSHASS-	166			
	: :				
Dd	102 MGEDDEDDEEEDEEDEE-----EDD-----EEGITGSNEDEDSSSKCSSV	145			
Oy	167 --DMTRPRRIGYTYVOQHIMGTDPRTILKDLP--TIIPPELMDMILMOIVINILISE	222			
	: :				
Dd	146 EPDMLR-----WLOREFYGRVPROYISIMEHFATGLAGDDDSVLAMYLAHLLINE	198			

[illegible]

	Query Match	30.3%	Score 1169;	DB 22;	Length 823;
	Best Local Similarity	38.8%;	Pred. No. 3,1e-87;		
	Matches 293;	Conservative 100;	Mismatches 206;	Indels 156;	Gaps 25;
OY	57 AAGCEAASAAAPALMTREA-----GAASAAREAPATAVAGDGDGSGSLRREPR-----	107			
	: : : : : : : : : : : : : : : : : : :				
Db	42 ASTSTEAEEAEATTTTEPATSELAGKANGCIKTKTLAAREDEIGANTLEHTKNPTKS	101			
OY	108 AADPDDEGEDEEDFAAAAAAAAAAICGYRDNLLITPGLLTNGFHSCEDDDDDTSHASS-	166			
	: : : : : : : : : : : : : : : : : : :				
Db	102 MGEDDDDEEEEEDEEEE-----BDD-----EEGTGTSNBEDSSSSSV	145			

```

QY 167 ---WTPRRRIGYTFYQOONHIMIGDPRHILANDLE--TIPPELDMITQIYNILISE 222
Db 146 EBDMLR-----WLOREFYTGVRPROVYLSIMHEFATGLAGDDDDSLMDYLAHLNE 198
QY 223 PPKRRKRDINTIEDAVNLLOECKKIIVLTGAGSVSGCIPDFRSRDSIYARLADFPDL 282
Db 159 PRRKRLASVTFDDIVILSKSQKIIYVLTGAGSVSGCIPDFRSTNGIYARLADFPDL 258
QY 283 PDPQAMFIEYFRKDPREFEFKAKIITYGQOPSLCHHFIALSDKEGKLRYNTQNTDPL 342
Db 259 PDPQAMFIEYFRKDPREFEFKAKIITYGQOPSLCHHFIALSDKEGKLRYNTQNTDPL 318
QY 343 EOVAIORILLOCHGSFATASCLICCYKYUDEAVRDNINOVVRCRPRP----- 391
Db 319 ERVAGIQIVIECHGSFSTASCCKCFKKNADALRADIDFQAIRPYCQOPKNEGVSVDASV 378
QY 392 --ADEPL-----ALMKPEIVFGSENLDEQFHRAMKYDKDEVDDLIVIGSSLKVRVALI 443
Db 379 AVTEELRLQIOLVENGIMKRPDIVFEGGELRDEYHTVATGDCDILLIVIGSSLKVRPAHI 438
QY 444 PSSITHEPQOLLINERPLPHLHFDVLELGDODVITNELCHRYLGE---YAKICCPVYLS 500
Db 439 PSSIPATVQOLLINERQOLHLHKFEDVLELGDSDVIITQICHLRSLDNDCCWRQOLCDBESVLT 498
QY 501 EITEKPPROKELVMSLPLPTPLHISIDSSSPER-----TVPODSSVIATLVQOA 551
Db 499 ESKELMP--PEHSHNHLH--HLHHRHCSSSEHQSQLDDPTDQSIKNSASADYILGSAG 554
QY 552 TNNVNDLEVSSESC-----VE-----EKPO-EVOT 576
Db 555 TCSD--SGESSFSGCGKRSATAAIAERIKDILIVELNETTALSCDRLGEGPQTIVES 613
QY 577 SRNV-----ENINVNPNPFKAVGSSTADKN--ERTSAEYTRKCKMPWRMLAKEQISKYLE 628
Db 614 YRHLSIDSSKDSGIEQCDNEATPSIVRSNLYOELKTYAASPLTPIPOGRKQOTAAERLQ 673
QY 629 -GNQYLFVPMNRYIEHGAEV-----YSDSEDDVLS--SCGSN---SDSGTC 670
Db 674 PGTFSHTNNYSYVPGAOQVEMDNDYSDODDEEERSHNRHSDLEGVNGHNYKDDDEDAC 733
QY 671 Q-----SPSLE-----EPLEDES 683
Db 734 DLNAVPLPLPSPLEAHIYTDIVYNGSNEPLPNSS 768

```

RESULT 11
 US-09-270-767-44940
 : Sequence 44940, Application US/09270767
 : GENERAL INFORMATION:
 : APPLICANT: Homburger et al.
 : TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 : FILE REFERENCE: File Reference: 7326-094
 : CURRENT APPLICATION NUMBER: US/09/270.767
 : CURRENT FILING DATE: 1999-03-17
 : NUMBER OF SEQ ID NOS: 62517
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 44940
 : LENGTH: 286
 : TYPE: PR1
 : ORGANISM: *Drosophila melanogaster*
 : FEATURE:
 : OTHER INFORMATION: Xaa means any amino acid
 : US-09-270-767-44940

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:43:12 ; Search time 76.21 Seconds
(without alignments)
109.926 Million cell updates/sec

Title: US-09-461-580A-4

Perfect score: 1288

Sequence: 1 VINILSEPPKRRKRDIINTI.....ALIPSSIPHEVPQILINREP 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468.5	36.4	257	21	Y73360
2	354.5	27.5	212	20	Y48540
3	138	10.7	205	19	W98718
4	97	7.5	689	21	Y49270
5	97	7.5	708	17	W57318
6	97	7.5	708	19	W57318
7	97	7.5	708	20	W86022
8	86	6.7	268	18	W20357
9	86	6.7	272	18	W21028
10	84	6.5	573	16	R77859
11	79.5	6.2	282	20	R37026
12	78.5	6.1	2366	17	R50511

13	77.5	6.0	2366	19	W68388
14	77	6.0	399	18	W14162
15	77	6.0	479	17	W04723
16	76.5	5.9	708	16	R71037
17	76	5.9	412	20	Y06396
18	76	5.9	516	20	Y07735
19	76	5.9	777	21	Y73385
20	76	5.9	2396	13	R29939
21	74	5.7	274	20	Y43931
22	74	5.7	297	20	W95686
23	74	5.7	945	20	Y35612
24	73.5	5.7	413	20	Y07057
25	73	5.7	142	19	W98856
26	72.5	5.6	620	16	R77275
27	72.5	5.6	620	18	W22225
28	72	5.6	956	14	R32356
29	72	5.6	3623	20	Y27020
30	71.5	5.6	696	20	Y13359
31	71.5	5.6	696	21	Y70671
32	71	5.5	469	19	W56793
33	71	5.5	707	21	Y95065
34	71	5.5	1702	14	R38095
35	70.5	5.5	245	20	Y14064
36	70.5	5.5	245	20	Y03651
37	70.5	5.5	245	20	W89436
38	70.5	5.5	245	20	W89433
39	70.5	5.5	541	16	R79581
40	70.5	5.5	805	20	Y06196
41	70.5	5.5	1367	13	R28041
42	70.5	5.5	1367	14	R44996
43	70.5	5.5	1367	14	R31377
44	70.5	5.5	1367	14	R37504
45	70.5	5.5	1367	15	R54046

ALIGNMENTS

RESULT 1

Y73360
ID Y73360 standard; Protein; 257 AA.

XX Y73360;

XX 24-FEB-2000 (first entry)

DT HTRM clone 2495790 protein sequence.

DE HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;
KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;
KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;
KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.

XX Homo sapiens.

OS WO9957144-A2.

PN 11-NOV-1999.

XX 04-MAY-1999; 99WO-US09935.

XX 05-MAY-1998; 98US-0084254.

PR 07-AUG-1998; 98US-0095827.

PR 02-OCT-1998; 98US-0102745.

XX (INCY-) INCYTE PHARM INC.

XX Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;

PI Gerstin EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;

XX WPI; 2000-052941/04.

DR N-PSDB; Z52445.

XX

PT New peptides useful for diagnosis, prevention and treatment of cancer
 XX and immune disorders

PS Claim 1; Page 120-121; 193pp; English.

XX Y73325-Y73389 are human transcriptional regulator molecule (HTRM) protein
 CC sequences. The HTRM protein and nucleotide sequences are useful for
 CC preventing or treating disorders associated with decreased expression or
 CC activity of HTRM which include cell proliferative disorders such as
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and
 CC leukaemia; immune disorders such as AIDS, Addison's disease, diabetes
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists
 CC of the HTRM polypeptides are useful for treating or preventing disorders
 CC associated with increased expression or activity of HTRMs. HTRM
 CC polypeptides, their immunogenic fragments or oligopeptides are useful for
 CC screening libraries of compounds in drug screening techniques.
 CC Polynucleotides encoding HTRM are useful for blocking the transcription
 CC of mRNA and regulating gene function by modulating the activity of HTRM.
 CC Vectors expressing HTRM or agonists can also be used to prevent or treat
 CC disorder associated with decreased HTRM expression. Antibodies which
 CC specifically bind HTRM and polynucleotides encoding HTRM are useful for
 CC diagnosing disorders associated with the expression of HTRM, particularly
 CC in assays that detect the expression of HTRM. Nucleotide sequences
 CC encoding HTRM may be useful to generate hybridization probes useful in
 CC mapping the naturally occurring genomic sequence and to detect
 CC differences in gene sequences among normal, carrier and affected
 CC individuals. Using diagnostic assays, cancer can be detected prior to the
 CC appearance of clinical symptoms and thereby progression of cancer can be
 CC prevented by aggressive treatment or preventive measures.

XX Sequence 257 AA;

Query Match 36.4%; Score 468.5; DB 21; Length 257;
 Best Local Similarity 43.9%; Pred. No. 8.6e-44;
 Matches 94; Conservative 47; Mismatches 58; Indels 15; Gaps 6;

QY 36 LTGAGVSCGIPDRS--RDGIYARLAVDFDLPDPQAMFDIEYFRKDPFFFKFAKIIY 94
 Db 1 mvgagistpsgipdrspgysnl--qqydlpyeafelfpffhnpkpfhtlatkely 58
 QY 95 PQQOPSLCHKFIALSDKEGKLLRNYTONIDTLEOVAGI--QRILQCHGSPATASCLICK 152
 Db 59 pgnypkpnvthyfrllhdkglllrytqnidglervsgipaskiveahgtfasatctvcq 118
 QY 153 YKVDCEAVRGDIFNOVPRCPRCPADEPLAIKPEIVFEGENLPEQF--HRAMKYDKDEV 210
 Db 119 rpfpgediradvmadrvrcpvc-----tgvvkpdvifgelpqgrfllh---vvdfpma 170
 QY 211 DLLIVIGSLKVRPVALIPSSIPHEVPQILINRE 244
 Db 171 dlilllgtsevepfasiteavrrsvprllnrd 204

RESULT 2

Y48540
 ID Y48540 standard; Protein; 212 AA.

XX AC Y48540;

DT 08-DEC-1999 (first entry)

XX Human breast tumour-associated protein 1.

XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
 KW treatment; tumour; cytostatic; medicament.

XX Homo sapiens.

XX DEJ9813839-A1.

XX 23-SEP-1999.

XX 20-MAR-1998; 98DE-1013839.
 XX 20-MAR-1998; 98DE-1013839.
 PR (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;

DR WPI; 1999-528981/45.

DR N-PSDB; 233613.

XX Human nucleic acid sequences and protein products from tumor breast
 PT tissue, useful for breast cancer therapy

PS Claim 25; 143; 188pp; German.

XX This invention describes novel human nucleic acid sequences from tumor
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer. Y48540-Y48617 represent protein fragments
 CC encoded by the expressed sequence tags described in the method of the
 CC invention.

XX Sequence 212 AA;

Query Match 27.5%; Score 354.5; DB 20; Length 212;
 Best Local Similarity 46.7%; Pred. No. 2.8e-31;
 Matches 70; Conservative 34; Mismatches 39; Indels 7; Gaps 4;

QY 5 LSEPPKRRKKRDKINTIEDAVKLLQ--ECKKIIVLTGAGVSCGIPDRS--RDGIYARLA 61

Db 56 lsllsqkerlldeitlegvarymqsercrrviclvgagistsagipdrspstglydnl- 114

QY 62 VDFDLPDPQAMFDIEYFRKDPFFFKFAKIIYFGQFQPSLCHKFIALSDKEGKLLRNYT 121

Db 115 -ekyhlpypeaifeislyfkkhpeffalakelypggfkptichyfmrlldkkglllrcyt 173

QY 122 QNIDTLEOVAGIQR--ILQCHGSPATASCL 149

Db 174 qnidtleriaagledeiveahgtfytshcv 203

RESULT 3

W98718
 ID W98718 standard; Protein; 205 AA.

XX AC W98718;

DT 31-MAR-1999 (first entry)

XX H. pylori GHPO 808 protein.

XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

[illegible]

```

CC      The invention provides an adhesion inhibitory composition of
CC      enteropathogenic Escherichia coli to cells. The composition contains a
CC      lactoferrin degraded matter. The composition is useful as a medicament,
CC      a drink and food or a feed. The composition is effective for prevention
CC      and/or treatment of infectious food poisoning caused by enteropathogenic
CC      E. coli.
CC      XX
SQ      Sequence      689 AA;

Query Match              7.5%; Score 97; DB 21; Length 689;
Best Local Similarity 23.6%; Pred No. 0.047;
Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9

OY      81 KDPREFEFMAKEIYDGPQPSLCHKFKIALSDK-----EGK----- 115
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      69 rdpylrlprvaeelgytksespqhyavavvkkgsmfgldqldgrrkschqldgrsawlp 128
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      116 --LLRNTYQUNIDTLEQVAVGIQRILLOCHGSEFATASCLIC---KRYVDCAVRNGDIFNQV 169
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      129 mgdliplyswteslepldg-----avakffasvcvpclrdqaypnlcqlckgsgenqca 182
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      170 PRCPRCPADPEP-----LAIMKPELVFEGEQLPQGFHAMKRYDDEVDL 212
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      183 -----cssarepyfgysgafkclqdgagdvatvketvtf--enle-----kadirqyei 229
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      213 LIVIGS-----SLKVRPVALIPS 230
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      230 lclnmstrapydafckechlaqvps 252
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT      5
ID      W09343
XX      W09343 standard; Protein; 708 AA.
AC      W09343;
XX      18-MAR-1997 (first entry)
DT
XX      Bovine lactoferrin.
DE
XX
XX      Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
KW      transferrin; bactericidal activity; prostate; expression system; primer;
KW      PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
XX      alpha-amylase; Aspergillus oryzae; nutrition; bovine.
XX
XX      Bos taurus.
OS
XX
XX      US5571691-A.
PN
XX      05-NOV-1996.
PD
XX      05-MAY-1989; 89US-0348270.
PE
XX      28-OCT-1993; 93US-0145681.
XX      05-MAY-1989; 89US-0348270.
PR      28-SEP-1989; 89US-0413880.
PR      24-APR-1992; 92US-0873304.
PR      27-OCT-1992; 92US-0967947.
XX
XX      (BAYU ) BAYLOR COLLEGE MEDICINE.
XX      PA
XX      Connely OM, Heaton DR, May GS, O'Malley BW;
XX      WPI; 1996-505400/50.
XX      DR      N-PSDB; T48031.
XX
XX      New DNA encoding human lactoferrin, its natural alleles and
PT      substitution analogues - useful e.g. for preventing iron deficiency
PT      and as antiviral/antimicrobial agent
XX
XX      Disclosure; Column 31-34; 92pp; English.
XX

```

CC This is the amino acid sequence of bovine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.
 CC Fragments of the protein are also known to have biological activity e.g.
 CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native
 CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
 CC peptide. The novel construct was then expressed in *Aspergillus oryzae*.
 CC The protein can be used for antibacterial and antiviral activities as
 CC well as an iron-carrying protein for nutritional or therapeutic
 CC applications.

XX Sequence 708 AA:

Query Match 7.5%; Score 97; DB 17; Length 708;

Best Local Similarity 23.6%; Pred. No. 0.049; Mismatches 57; Indels 72; Gaps 9;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

OY 81 KDRPPEFFAKETIYGCGQPSLCHKFIASLSDK-----EGK----- 115
 DB 88 rdpkrlrpvaeeiygtkspqthyyavaavvkgsnfgldqgkrkschtlgirsagwllp 147
 OY 116 --LLNNTQNTIDLEQVAGIQRIILQCHGSFATASCLIC---KYVDEAVAGDIFNQV 169
 DB 148 mglitpylswtestlepldg-----avakffsascvpcldqaypnclqckgegenqca 201
 OY 170 PRCPRCPADP-----LAIMKPEIVFEGENTLPEQFHRAMKYDDEVLD 212
 DB 202 -----csarepyfygsgafkclqdgagdvafyketvfe--enlpe-----kadtqyel 248
 OY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 lclnstrapvdafkechlaqyps 271

RESULT 6

W57318 ID W57318 standard; Protein; 708 AA.

AC W57318;

XX 10-AUG-1998 (first entry)

DE Bovine lactoferrin.

XX Bovine lactoferrin; recombinant; therapeutic; nutritional; iron;

KW Fe binding site; bacteria; bactericidal; milk.

XX Bos taurus.

OS US5766939-A.

PN US5766939-A.

XX 16-JUN-1998.

PD 30-MAY-1995; 95US-0453703.

XX 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PR 30-MAY-1995; 95US-0453703.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Connely OM, Headon DR, May GS, O'Malley BW.

XX WPI; 1998-361705/31.

XX N-PSDB; V30771.

XX Plasmids containing human lactoferrin DNA - for recombinant

XX production of the enzyme, especially fragments having bactericidal

PS Example 13; Fig 14; 92pp; English.

XX The present sequence represents bovine lactoferrin. A plasmid has been
 CC developed which is suitable for the expression of a human lactoferrin,
 CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
 CC host cell. The plasmid comprises a DNA sequence encoding a naturally
 CC occurring human lactoferrin protein or an iron-binding lobe of
 CC lactoferrin and further comprises transcriptional and translational
 CC regulatory elements capable of regulating the expression of the
 CC lactoferrin-encoding DNA sequence in the transformed host cell. The
 CC plasmid is useful for producing recombinant human lactoferrin proteins
 CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
 CC in milk and other secretory fluids. It is involved in iron transfer
 CC and delivery in mammals. It has been implicated as a resistance
 CC factor in suckled new born infants against enteric infections: the
 CC iron binding activity thought to be responsible for bactericidal action
 CC as the bound iron deprives microorganisms of the metal, and interferes
 CC with their growth and reproduction. The protein can be used for
 CC nutritional and therapeutic applications. It has not been previously
 CC possible to efficiently produce lactoferrin recombinantly.

XX Sequence 708 AA:

Query Match 7.5%; Score 97; DB 19; Length 708;

Best Local Similarity 23.6%; Pred. No. 0.049; Mismatches 57; Indels 72; Gaps 9;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

OY 81 KDRPPEFFAKETIYGCGQPSLCHKFIASLSDK-----EGK----- 115
 DB 88 rdpkrlrpvaeeiygtkspqthyyavaavvkgsnfgldqgkrkschtlgirsagwllp 147
 OY 116 --LLNNTQNTIDLEQVAGIQRIILQCHGSFATASCLIC---KYVDEAVAGDIFNQV 169
 DB 148 mglitpylswtestlepldg-----avakffsascvpcldqaypnclqckgegenqca 201
 OY 170 PRCPRCPADP-----LAIMKPEIVFEGENTLPEQFHRAMKYDDEVLD 212
 DB 202 -----csarepyfygsgafkclqdgagdvafyketvfe--enlpe-----kadtqyel 248
 OY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 lclnstrapvdafkechlaqyps 271

RESULT 7

W86022 ID W86022 standard; Protein; 708 AA.

AC W86022;

XX 22-FEB-1999 (first entry)

DE Bovine lactoferrin (hLF) polypeptide.

XX Lactoferrin; hLF; recombinant; iron-binding; bovine; porcine; eyedrop;

KW *Aspergillus*; animal food; therapeutic additive; iron transport; human;

KW vitrucidal; bactericidal; contact lens; skin care; eardrop; mouthwash;

KW chewing gum; toothpaste; preservative; anti-infection; nutrition.

XX Bos taurus.

OS US5849881-A.

PN US5849881-A.

XX 15-DEC-1998.

XX 30-MAY-1995; 95US-0456106.

XX 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PR 30-MAY-1995; 95US-0456106.

XX (CONN/) CONNELLY O M.
 PA (HEAD/) HEADON D R.
 PA (MAYG/) MAY G S
 PA (OMAL/) O'MALLEY B W.
 XX
 PI Conneely OM, Headon DR, May GS, O'Malley BW.
 XX
 DR WPI: 1999-069808/06.
 DR N-PSDB: V80376.
 XX
 PT Recombinant lactoferrin - and lactoferrin fragments
 XX
 PS Example 13; Columns 33-38; 93pp; English.
 XX
 CC This represents a bovine lactoferrin polypeptide. The invention relates
 CC to production of recombinant lactoferrin using cDNA sequences of human,
 CC bovine and porcine lactoferrin. Regions of the cDNA such as the iron-
 CC binding sites can be used to produce recombinant lactoferrin, especially
 CC human lactoferrin (hLF) polypeptide product. Aspergillus host cell
 CC transformed with an expression vector comprising a hLF DNA sequence, a
 CC promoter, a translation initiation sequence and transcription and
 CC translation termination sequences can be used for the production of the
 CC recombinant hLF polypeptide. Recombinant lactoferrin polypeptides can be
 CC used in human and animal foods, as therapeutic additives to enhance iron
 CC transport and delivery, and for their virucidal and bactericidal
 CC properties, as additive for eyedrops, contact lens care solutions, skin
 CC care products, eardrops, mouthwashes, chewing gum and toothpaste, as
 CC preservatives for the above products, as therapeutic anti-infection
 CC agents, as nutritional supplements and as sources of amino acids.
 XX
 SQ Sequence 708 AA;

Query Match 7.5%; Score 97; DB 20; Length 708;
 Best Local Similarity 23.6%; Pred. No. 0.049;
 Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

OY 81 KDPKPFKFKKEIYPGQFQPSLCKHFTALSDK-----EGK----- 115
 DB 88 rdpkllpvaaellygtespqthyaavavvkgsnfgldqgqkschtg1gragwllp 147
 OY 116 --LLRNTQNIIDLEQVAGIQRILQCHGSFATASCLIC---KKVDCNAVREGDIFNOVV 169
 DB 148 mglirp1jswtesleplqg-----avakfiascvcpcidqgypnlcqlkgegenqca 201
 OY 170 PRCPRCPADEP-----LAIKPEIYFEGENLPEQFHRAMKYDKDEVDL 212
 DB 202 -----csrepyfgysgafkclgdgagdvafvketvfe--enlpe-----kadrdygel 248
 OY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 lclnnsrapvdafkchlaqyps 271

RESULT 8

W20357
 ID W20357 standard; Protein; 268 AA.
 XX
 AC W20357;
 XX
 XX
 DT 11-JUL-1997 (first entry)
 XX
 DE H. pylori transmembrane protein 26351567.aa.
 XX
 XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 PN MO9640893-AA.
 XX

PD 19-DEC-1996.
 XX
 PE 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglinth OT, Smith D, Mellgaard BL;
 XX
 DR WPI: 1997-052306/05.
 DR N-PSDB: T67549.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 548; 1481pp; English.

CC This sequence is a H. pylori transmembrane protein likely to
 CC contain at least two membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 268 AA;

Query Match 6.7%; Score 86; DB 18; Length 268;
 Best Local Similarity 23.6%; Pred. No. 0.21;
 Matches 34; Conservative 21; Mismatches 49; Indels 40; Gaps 6;

OY 120 YTONIDTLEQVAGIQRILQCHGSFATASCLICK---KVD-----C 157
 DB 30 yegeik--ekavvnvsg-----gkiftsyilfqayegiknkidltinqvknndetksvnenl 82
 OY 158 EAVRGDI--FNQVPRCPRCPADEPLAIKPEIYFEGENLPEQFHRAMKYDKDEVDLIV 215
 DB 83 ektgkldlddlnalvqklnlpkdfnkclikpqsflfnynta-----nedeknrlvi 133
 OY 216 IGSSLKVRPALIPSSIPHEVPOI 239
 DB 134 lasrlsqketqppslsknsvshl 157

RESULT 9

W21028
 ID W21028 standard; Protein; 272 AA.
 XX
 AC W21028;
 XX
 XX
 DT 23-JUL-1997 (first entry)
 XX
 DE H. pylori transmembrane protein, hp6p10723orf13.
 XX
 XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KM identification; binding compound; bacterium; life cycle; activator;
 KM bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KM diagnosis.
 XX
 OS Helicobacter pylori.
 XX

OS Clostridium difficile.
 XX
 PN W03808540-A1.
 XX
 PD 05-MAR-1998.
 XX
 XX 28-AUG-1997; 97WO-US15394.
 XX
 XX 28-AUG-1996; 96US-0704159.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Thalley BS, Williams JA;
 XX
 DR WPI: 1998-230234/20.
 DR N-PSDB: V30561.
 XX
 PT Host cell containing recombinant expression vector encoding
 PT Clostridium botulinum type B or E toxin - useful to treat humans
 PT and other animals at risk of intoxication with clostridial toxin
 XX
 PS Example 18; Page 241-249; 428pp; English.
 XX
 CC This is the amino acid sequence of Clostridium difficile toxin B,
 CC deduced from the coding region (see V30561) of the toxin B gene.
 CC Fragments of the toxin B gene have been cloned into various
 CC prokaryotic expression systems, and assessed for the ability to
 CC express recombinant toxin B protein in E. coli. It would be
 CC advantageous to use simple and inexpensive prokaryotic expression
 CC systems to produce and purify high levels of recombinant toxin B
 CC for immunisation purposes. The invention specifically relates to
 CC recombinant proteins derived from Clostridium botulinum toxins
 CC (see W68399-000) and their use as immunogens for the production of
 CC vaccines and antitoxins.
 CC
 XX Sequence 2366 AA:
 SQ
 Query Match 6.0%; Score 77.5; DB 19; Length 2366;
 Best Local Similarity 17.8%; Pred. No. 38;
 Matches 54; Conservative 50; Mismatches 125; Indels 75; Gaps 8;
 QY 2 INITLS-----EPPKRRKKRDINTIEDAVKLLDECKKIIVLTGAGVSCGIPDF 50
 DB 836 IIVININLTQVVEERLEEAKNLTSGSINYIKDEFKLE-----SISDALCOL 872
 QY 51 RSRDGIYARLAVDPDLPDPPQAMFDEYRKDRP--FFKFAKEIYQGFQSLCHKFIA 108
 DB 873 KQGNLEDSHISFIDISCTDEGFIIRINKETGEISIVETKTLF-seyanhiteelsk 931
 QY 109 LSDK-----EGKLLR---NYTONIDTLEQVAGIORIILQCHGSFATASCLICKYKVDCEA 159
 DB 932 Ikgltftvnglrvkvkndtthevntlnaeflgsllleyskseslnlsvamkvqyva 991
 QY 160 -----VRGDIENOVVPRCPRCPADPEFLAIKPEIYFFGNNLDEQFHR 201
 DB 992 qlfstgltltlaakvelstalcetldlptlseglptlatitldgvsigaikelset 1051
 QY 202 AMKYDKDEVD-----LLIYIGSSLKVRPALVLPSSIPHEVQPL 240
 DB 1052 sopllrtgeaklgimavnlttaltaltsslgiasgfsillvplaglsagipslvnnel 1111
 QY 241 INRE 244
 DB 1112 vlrld 1115
 RESULT 14
 ID W14162
 AC W14162 standard; Protein: 399 AA.
 XX W14162:
 XX

DT 24-JUN-1997 (first entry)
 XX
 DE R. rhodochrous protein encoded by putative open reading frame.
 XX
 KM Rhodococcus rhodochrous; amidase; probe; amplification; primer; PCR;
 KM polymerase chain reaction; phage; open reading frame; nitrile hydratase;
 KM enzyme; genetic engineering.
 XX
 OS Rhodococcus rhodochrous.
 XX
 PN JP09009973-A.
 XX
 PD 14-JAN-1997.
 XX
 XX 27-JUN-1995; 95JP-0184934.
 XX
 PR 27-JUN-1995; 95JP-0184934.
 XX
 PA (CHCC) CHISSO CORP.
 XX
 DR WPI: 1997-126430/12.
 DR N-PSDB: T62387.
 XX
 XX Rhodococcus rhodochrous nitrile hydratase and amidase genes -
 PT useful for the industrial preparation of useful cpds.
 PT
 PS Examples; Page 13-15; 16pp; Japanese.
 XX
 CC This is the amino acid sequence of a protein of unknown function encoded
 CC by a fourth open reading frame (ORF) contained in the fragment (T62385)
 CC isolated from Rhodococcus rhodochrous strain IF015564 which contains the
 CC ORFs for the amidase and nitrile hydratase alpha and beta subunits. The
 CC fourth ORF is hypothetical and is located downstream of a terminator
 CC sequence at the 3' end of the nitrile hydratase beta subunit coding
 CC sequence. The fragment containing the ORF was isolated by screening a
 CC genomic R. rhodochrous DNA library in EMBL3 vector, using a probe
 CC amplified by the primers T62388-9. The primers amplified a 0.7 kb
 CC fragment of the amidase gene. The library screen isolated a phage
 CC containing a 15 kb insert comprising the above mentioned ORFs.
 CC The enzymes, prepared by genetic engineering methods, can be used in the
 CC industrial production of useful compounds.
 CC
 XX Sequence 399 AA:
 SQ
 Query Match 6.0%; Score 77; DB 18; Length 399;
 Best Local Similarity 25.4%; Pred. No. 3.6;
 Matches 36; Conservative 20; Mismatches 46; Indels 40; Gaps 8;
 QY 98 FQPSLCHKRIALSDREGKLLR---NYTONIDTLEQVAGIORIILQCHGSFATASCLICKYKV 155
 DB 264 fhpgtrlhdfllsewengklrlakgywnagrtfel-----gslsqaghlrlrhyv 313
 QY 156 DCEAVRGDIENOVVPRCPRCPAD-----EPLAIKPEIYFFGNNL-PEQFHRA 202
 DB 314 -----grwmn-flpr-dewpadgyrtdgldkweeypdcrgelvfpgaldpsrlhr- 364
 QY 203 MKYDKDEVLLIVIGSSLKVRP 224
 DB 365 -----eldacilttaeie19p 380
 RESULT 15
 ID W04723
 AC W04723 standard; Protein: 479 AA.
 XX W04723:
 XX
 DT 06-FEB-1997 (first entry)
 XX
 DE Aromatic acyl transferase of Gentiana triflora var. japonica.
 XX
 XX Aromatic acyl transferase; transformation; anthocyanin pigment;
 XX

KW plants; acylation; colour; tone; colouration; colour change;
 KW Gentiana triflora; Petunia hybrida; Petilla oclimoides;
 KW Senecio cruentus; Lavandula angustifolia.

OS Gentiana triflora var. japonica (Clone pcAT106).

XX WO9625500-A1.

XX 22-AUG-1996.

XX 16-FEB-1996; 96WO-JP00348.

XX 30-JAN-1996; 96JP-0046534.

PR 17-FEB-1995; 95JP-0067159.

PR 29-JUN-1995; 95JP-0196915.

XX (SUNR) SUNTORY LTD.

XX Ashikari T, Fujiwara H, Fukui Y, Kusumi T, Mizutani M;
 PI Nakao M, Tanaka Y, Yonekura K;

XX WPI, 1996-393401/39.

DR N-PSDB; T37309.

PT DNA coding for aromatic acyl transferase - for transforming plants
 PT which produce anthocyanin pigments and thus altering colour tone,
 PT e.g. of flowers

XX Claim 4; Page 57-61; 94pp; Japanese.

XX Vectors containing DNA fragments encoding proteins of plant origin
 CC with aromatic acyl transferase activity may be used to transform

CC plants which produce anthocyanin pigments. The aromatic acyl
 CC transferase acylates the pigments in the flower resulting in colour

CC tone changes and allowing new colourations to be produced. Six
 CC specific DNA sequences encoding aromatic acyl transferase from

CC different plants are described in T37308-T37313.

XX Sequence 479 AA;

Query Match 6.0%; Score 77; DB 17; Length 479;

Best Local Similarity 20.8%; Pred. No. 4.6; Mismatches 71; Indels 56; Gaps 9;

DB 70 PQAMPDIYFRKDPPEFKAEIYPCGQPSLC--HKETALSDKEG-----KLLR 118
 29 plsfddmrwlisdaenhlfyfrhp-----cpnskfissiksalsivlkhlplar 81

QY 119 NYTQNTDTLEQVAGIQRILOCHGSFAFASCLICKYKVDCAVRGD-----IRQVVPRC 172
 82 nlwpydssdrmpelr---ykkgsdvslltlessmdfyldqrdyktfndilpq 136

QY 173 PR---CPADEPLAIKMPKEIYF-----GENTPECFHRAKRY-----DKDEVDL 212
 137 pepivsgdevlpfalqvfvfntgicigrnlhqvlgdassfhnfklwlvdkn--- 193

QY 213 LIVIGSSLKVRVVALIP 229
 194 ---gdsllkftlplsslp 206

Search completed: February 16, 2001, 10:43:16
 Job time: 98 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:45:24 ; Search time 80.58 Seconds
(without alignments)
206,449 Million cell updates/sec

Title: US-09-461-580A-4

Perfect score: 1288

Sequence: 1 VINILSEPPKRRKKRDINTI.....ALIPSPHEVPQILINREP 245

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR66:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	731.5	56.8	607	2 T24172	hypothetical prote
2	551	42.8	670	2 S36616	regulatory protein
3	535.5	41.6	503	2 S59698	HST1 protein - yea
4	518	40.2	562	1 R6852	regulatory protein
5	515	40.0	471	2 T39571	probable regulator
6	470	36.5	381	2 JC4639	silent information
7	466	36.2	332	2 T40929	transcription regu
8	455	36.1	331	2 T18247	transcription regu
9	418	32.5	357	2 S59678	HST2 protein - yea
10	346	26.9	246	2 A72370	regulatory protein
11	331.5	25.7	247	2 H69827	conserved hypotet
12	318	24.7	253	2 H69263	transcription regu
13	315	24.5	447	2 S54631	HST3 protein - yea
14	287.5	22.3	245	2 C69459	transcription regu
15	282.5	21.9	249	2 F71085	hypothetical prote
16	277	21.5	247	2 D72562	hypothetical prote
17	267.5	20.8	250	2 C75101	transcription regu
18	252.5	19.6	250	2 A83506	probable cobalamin
19	246	19.1	415	2 T50106	hst4p [imported] -
20	237.5	18.4	239	2 C70486	conserved hypotet
21	229	17.8	370	2 S52699	hypothetical prote
22	220	17.1	287	2 T22324	hypothetical prote
23	210	16.3	259	2 E82191	probable nicotinat
24	206	16.0	279	2 E64856	hypothetical prote
25	201.5	15.6	237	2 H70554	hypothetical prote
26	200.5	15.3	256	2 A83495	hypothetical prote
27	197.5	15.3	234	2 A71838	hypothetical prote
28	190	14.8	287	2 T22325	hypothetical prote
29	187	14.5	233	2 G81307	probable transfera

30	185.5	14.4	246	2 G75570	conserved hypotet
31	150.5	11.7	190	2 T35951	probable SIR2 faml
32	139.5	10.8	1095	2 T25520	hypothetical prote
33	138	10.7	205	2 C64677	conserved hypotet
34	135.5	10.5	118	2 T46348	hypothetical prote
35	103	8.0	87	2 T34648	probable SIR2 faml
36	97	7.5	708	1 TFBOL	lactotransferin p
37	89	6.9	259	1 B69407	Mu0188 homolog - A
38	88	6.8	320	2 E69264	pyruvate formate-1
39	86	6.7	221	2 G72322	glutaredoxin-relat
40	86	6.7	268	2 C71872	hypothetical prote
41	85.5	6.6	482	2 H69392	4-hydroxybutyrate
42	84.5	6.6	527	2 S75829	hypothetical prote
43	83.5	6.5	1905	2 I51553	plexin - African C
44	83	6.4	355	2 E64445	polyterredoxin 4x2
45	83	6.4	905	2 T38980	hypothetical prote

ALIGNMENTS

RESULT 1
T24172
hypothetical protein R11A8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24172
R:Cummings, P.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19849
A:Accession: T24172
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <WT>
A:Cross-references: EMBL:270310; PIDN:CAA94364.1; GSPDB:GN00022; CESP:R11A8.4
C:Experimental source: clone R11A8
C:Genetics:
A:Gene: CESP:R11A8.4
A:Map position: 4
A:Introns: 18/3; 68/3; 111/3; 206/3; 500/3; 565/3

Query Match 56.8%; Score 731.5; DB 2; Length 607;
Best Local Similarity 57.0%; Pred. No. 6.4e-38;
Matches 138; Conservative 39; Mismatches 60; Indels 5; Gaps 1;
QY 3 NILSEPPKRRKKRDINTIEDAVKLOECKRIIVTGAGVSVCIGIPDFRSRDGIYRLAV 62
Db 118 DLERAVRKQKLTNYNSLADAVELFKTKKHLVLTGAGVSVCIGIPDFRSKDGIYRLRS 177
QY 63 DEPDLPDQAMFIDIEYFRKDPPEFKAKETIYQGFQPSLCKHPIALSDEKLLRNVTQ 122
Db 178 EFPDLPDPTAMFDIYFRFNAPYNFARIEFPQFVPSVSHRIKELETSGRLLRNVTQ 237
QY 123 NIDPLEVAGIQRLOCHGSFATASCLIKYKVCFAVRGDFINOVYPRPCRADEPLA 182
Db 238 NIDPLEHQGTGIRVVECHSFSCTCRGQKYGNGEIVRELVAMRAHCKRCE----G 292
QY 183 IMKEIYFEGENLEPEOFHRAKRYKDEVDLLIVIGSSILKYPVALIPSSIPHEVPQILIN 242
Db 293 VIRKNIYFEGEDLGRFHQHTEDKHKYVDLIVIGSSILKYPVALIHCYDKNPQILIN 352
QY 243 RE 244
Db 353 RE 354
RESULT 2
S36616
regulatory protein SIR2 - yeast (Kluyveromyces marxianus var. lactis)
N:Alternate names: silent information regulator 2
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999

C:Accession: A56048; S36616
R:Chen, X.J.; Clark-Walker, G.D.
Mol. Cell. Biol. 14, 4501-4508, 1994
A:Title: sir2 mutants of *Kluyveromyces fragilis* are hypersensitive to DNA-targeting drugs
A:Reference number: A56048; MUID:94277035
A:Accession: A56048
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-670 <CH2>
A:Cross-references: EMBL:X74569; NID:G397349; PIDN:CA52661.1; PID:G397350
C:superfamily: regulatory protein sir2
C:keywords: DNA binding; nucleus; transcription regulation

[illegible]

RESULT 3

S59698

HST1 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein O1161; protein YOL0686

C:Species: *Saccharomyces cerevisiae*

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Jun-2000

A:Accession: S59698; S62108; S66761; S70376

R:Derbyshire, M.K.; Weinstein, K.G.; Strathern, J.N.

submitted to the EMBL Data Library, September 1995

A:Description: HST1, a new member of the SIR2 family of genes.

A:Reference number: S59698

A:Accession: S59698

A:Molecule type: DNA

A:Residues: 1-503 <DER>

A:Cross-references: EMBL:L47120; NID:9972892; PIDN:AA38430.1; PID:9972893

R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D.

submitted to the EMBL Data Library, October 1995

A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in silencing of repetitive DNA sequences.

A:Reference number: S62108

A:Accession: S62108

A:Molecule type: DNA

A:Residues: 1-503 <BRA>

A:Cross-references: EMBL:U39041; NID:91055019; PIDN:AAA81033.1; PID:91055020

R:Alexandriaki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66756

A:Accession: S66761

A:Molecule type: DNA

A:Residues: 1-503 <ALE>

A:Cross-references: EMBL:Z74810; NID:91419890; PIDN:CAA9078.1; PID:91419891; MIPS:YOL00682

A:Experimental source: strain S288C

R:Derbyshire, M.K.; Weinstein, K.G.; Strathern, J.N.

Yeast 12, 631-640, 1996

A:Title: HST1, a new member of the SIR2 family of genes.

A:Reference number: S70376; MUID:96405912
A:Accession: S70376
A:Molecule type: DNA
A:Residues: 1-503 <DEM>
A:Cross-references: EMBL:LA7120; NID:g972892; PIDD:AA838430.1; PID:g972893
C:Genetics:
A:Gene: SGD:HST1
A:Cross-references: SGD:S0005429; MIPS:YOL068c
A:Map position: 15L
C:Superfamily: regulatory protein SIR2
C:Keywords: transmembrane protein
F:1-17/Domain: transmembrane #status predicted <TM1>
F:402-418/Domain: transmembrane #status predicted <TM2>

[illegible]

RESULT 4
 REGSVS2
 regulatory protein SIR2 - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein D2714; protein YDL042c; silent information regulator 2
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
 C:Accession: S05891; S67575
 R:Shore, D.; Squire, M.; Naamlyth, K.A.
 A>Title: Characterization of two genes required for the position-effect control of yeast
 A:Reference number: S05891; MUID:85126876
 A:Accession: S05891
 A:Molecule type: DNA
 A:Residues: 1-562 <SR0>
 A:Cross-references: EMBL:X01419; NID:94469; PIDD:CAA25667.1; PID:g4470
 R:Paulin, L.; Saren, A.M.; Laamanen, P.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67560
 A:Accession: S67575
 A:Molecule type: DNA
 A:Residues: 1-562 <PAU>
 A:Cross-references: EMBL:Z74090; NID:g1431026; PIDD:CAA98600.1; PID:g1431027; GSPDB:G-
 A:Experimental source: Strain S288C
 C:Genetics:
 A:Gene: SGD:SIR2; MIPS:YDL042c
 A:Cross-references: SGD:S0002200; MIPS:YDL042c
 A:Map position: 4L
 C:Superfamily: regulatory protein SIR2
 C:Keywords: DNA binding; transcription regulation

Query Match	Score	DB 1	Length
Best Local Similarity	Pred. No.	9.2e-39	
40.28;			
39.18;			

Matches 117; Conservative 44; Mismatches 80; Indels 58; Gaps 7;

Query 1 VINLSEPPKRRK-----KDIN-----TIEDAVKLLQCKKIIVLTG 38
 Db 205 VTNN--EDPLAKKOTVRLIKLQRAINKVLCITRLRLSNFEFTIDHTQILHAKRILVLTG 262

Query 39 AGVSVSCGIDPFRSRDGIYARLAVDFPDLPPQAMFDEYFRKDPPEFRFAKEIYQGF 98
 Db 263 AGVSTSLGIDPFRSSEGFYSK--IKHLGLDPQDVFNNTIMHDPVSYYNANVLPPEK 320

Query 99 QPSLCHKFIALSDKEGKLLRNVTQNTIDTLEQVAGI--QRILOCHGSFATASCLICKYKD 156
 Db 321 IYSLHSGFIKMLQMKKGLRNVTQNTIDTLEQVAGI--QRILOCHGSFATASCLICKYKD 380

Query 157 CEAVRGDIFNOVPRC-----PADEP-----LAIMKP 186
 Db 381 GERLEFNKIRNLEPLCPYCKKRRREYEGYNNKGVAAAGSMSERPPYILNSYGLKP 440

Query 187 EIVFEGENLPEQFHRANKYDKDEVLLIVIGSLKVRVALIPSSIPHEVQIILNREP 245
 Db 441 DITFEGALPKFKHKSITREDILECDLLICISTSLKAVSETVNMVPSHVQVILNRP 499

RESULT 5
 T39571
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T39571
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.
 submitted to the EMBL Data Library, March 1999
 A:Reference number: Z21864
 A:Accession: T39571
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-471 <MOO>
 A:Cross-references: EMBL:AL035637; PIDN:CA83511.2; GSPDB:GN00067; SPDB:SPBC16D10.07C
 A:Experimental source: strain 972n-; cosmid c16D10
 C:Genetics:
 A:Gene: SPDB:SPBC16D10.07C
 A:Map position: 2
 A:Introns: 53/1; 106/3; 117/2; 263/1

Query Match 40.0%; Score 515; DB 2; Length 471;
 Best Local Similarity 43.6%; Pred. No. 1.4e-38;
 Matches 119; Conservative 39; Mismatches 77; Indels 38; Gaps 7;

Query 7 EPPRRKKRNDINTEDAVKLLQCKKIIVLTGAGVSVSCGIDPFRSRDGIYARLAVDFP 66
 Db 129 EVARRIKLPHENTEFEDVNNLKAKNNVVLGAGISTSLGLDRSDGFAARLARH--G 186

Query 67 LPDPQAMFDEYFRKDPPEFRFAKEIY--GQFQPSLCHKFIALSDKEGKLLRNVTQNI 124
 Db 187 LSESEMFHDITFRENPEIETFAFDLLEFNHSPS--HAFILLEKKNKLSLFTQNI 244

Query 125 DTLEQVAGI--QRILOCHGSFATASCLICKYKDCEAVRGDIFNOVPRC----- 175
 Db 245 DNLEKKTGLSDNKLIOCHGSFATATCIKCKHRYVGSSELYEDIRNRVSYCNCKGKPLKL 304

Query 176 -----PADEP-----IMKPEIVFEGENLPEQ--FHRANKYDKDEVLL 212
 Db 305 RRVGQNKKEKHYSDDGSESEDDLAQPGIKMKPDITFEGALPDSFKNVSGSELEETDL 364

Query 213 LIVIGSLKVRVALIPSSIPHEVQIILNREP 245
 Db 365 LICIGTSLKAVAPVSELISVIPPTTPQIYISKRP 397

RESULT 6
 JC4639
 C:Species: Leishmania major
 C>Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 10-Sep-1997
 C:Accession: JC4639; PC4152
 R:Tahiaoui, B.; Taibi, A.; Ouasssi, A.
 Gene 169, 115-118, 1996
 A:Title: A Leishmania major protein with extensive homology to silent information reg
 A:Reference number: JC4639; MUID:96186914
 A:Accession: JC4639
 A:Molecule type: mRNA
 A:Residues: 1-381 <YAH>
 A:Cross-references: GB:L40331; NID:g1203986; PID:g1203987
 A:Accession: PC4152
 A:Molecule type: protein
 A:Residues: 58-76 <YAH>
 C:Genetics:
 A:Gene: LmsIR2rp
 C:Keywords: zinc finger
 F:270-310/Region: serine-rich
 F:152,155,176,179/Binding site: zinc (Cys) #status predicted

Query Match 36.5%; Score 470; DB 2; Length 381;
 Best Local Similarity 41.3%; Pred. No. 1.2e-34;
 Matches 102; Conservative 45; Mismatches 76; Indels 24; Gaps 7;

Query 5 LSEPPKRRKNDINTEDAVKLLQCKKIIVLTGAGVSVSCGIDPFRSRD-GIYARLA 61
 Db 14 LGEP-----TVEGLARYIREKDVRLIVLGAAVAAAGIPDRSSDGTGYARKLG 63

Query 62 VDFPDLPPQAMFDEYFRKDPPEFRFAKE--TYGQFQPSLCHKFIALSDKEGKLLRN 119
 Db 64 K--YNLDDPTAFSTILIREKPEIFYSTARELNIMPFGHFOPTAVNHFRLQDDEGRLLRC 121

Query 120 YTONIDTLEQVAGI--QRILOCHGSFATASCLICKYKDCEAVRGDIFNOVPRCPA 177
 Db 122 CTQNDIGLEKAGVSPELLVEAHGSFAAACIECHTPTSIQNTLEAMSGTVSCSTG- 180

Query 178 DEPLAIMEIYFEGENLPEQFHRANKYDKDEVLLIVIGSLKVRVALIPSSIPHEVP 237
 Db 181 ---GIVKPNVYFEGENLPDAFFDLHHDAPIAELVLIIGTSMOVHPFALLPCVAPKSP 236

Query 238 QIILNRE 244
 Db 237 RVLMNRE 243

RESULT 7
 T40929
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 C:Accession: T40929
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Aert, R.; Volckaert, G.
 submitted to the EMBL Data Library, October 1999
 A:Reference number: Z21958
 A:Accession: T40929
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-332 <LYN>
 A:Cross-references: EMBL:AL121807; PIDN:CA858129.1; GSPDB:GN00068; SPDB:SPCC132.02
 A:Experimental source: strain 972n-; cosmid c132
 C:Genetics:
 A:Gene: SPDB:SPCC132.02
 A:Map position: 3
 A:Introns: 27/1; 152/3

Query Match 36.2%; Score 466; DB 2; Length 332;
 Best Local Similarity 42.4%; Pred. No. 2.3e-34;
 Matches 98; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

Query 20 IEDAVKLLQCKKIIVLTGAGVSVSCGIDPFRSRD-GIYARLAVDFPDLPPQAMFDI 76
 Db 100 IEDAVKLLQCKKIIVLTGAGVSVSCGIDPFRSRD-GIYARLAVDFPDLPPQAMFDI 76

Db 15 LEKVASLIKEGVKKICVWVGAGISTAGIDPFSPETGTYNNL--QRFNLPYAEAVEDL 72
 Qy 77 EYFRKDPPEFFAKETIYPGQFSLCHKFIALSDKEGLLNRYTONIDTLEQVAGI--Q 134
 Db 73 SYFRKNRPFFYLAEHMEKRYPTHTYFIRLHDRLKLOCYTONIDTLERLAGVPRK 132
 Qy 135 RILQCHGSFATASCLICKKYKVDCEAVRGDIFNOVVRPCRPCRADEPLAIKPEIYFEGEN 194
 Db 133 ALIEAHGSFOYSRCIECYEMAEYEVACIMOKQVPCNSCK----GLIKPMIYFYGEG 187
 Qy 199 LPEQHRANKYDKDEVLLIVIGSSLKVRPALIISSTIHEVPQILINER 245
 Db 188 LPMREFHEMEKDTKVDMAVLTGISTLVHPFADLPPEIYVPMKORVILINER 238

RESULT 8
 118247
 transcription regulatory protein - yeast (Candida albicans)
 C:Species: Candida albicans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18247
 R:Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z18831
 A:Accession: T18247
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-331 <BAR>
 A:Cross-references: EMBL:AL033503; NID:e1341066; PID:e1341074; PIDN:CAA22018.1
 C:Genetics:
 A>Note: Ca49C4.08c

Query Match 36.1%; Score 465; DB 2; Length 331;
 Best Local Similarity 40.8%; Pred. No. 2.9e-34;
 Matches 95; Conservative 53; Mismatches 71; Indels 14; Gaps 5;

Qy 19 TIEDAVKLLQEC---KKIIVLTGAGVSVSCGIPDFRSND-GIYARLAVDPDLPPQAM 73
 Db 3 SLDLILKPAVAELVKNKGKVTFFNGAGISTGAGIDPFRSPDGLYANLAK--LNLPRFAEAV 60
 Qy 74 FIEIEFRKDPPEFFAKETIYPGQFSLCHKFIALSDKEGLLNRYTONIDTLEQVAGI 133
 Db 61 FDIIEFKEDPKPEPYTLAEELYPGNFAPKFNHFIKLQDGSILKRYTONIDTLERLAGV 120
 Qy 134 Q-RILQCHGSFATASCLICKKYKVDCEAVRGDIFNOVVRPCRPCRADEPLAIKPEIYF 191
 Db 121 EKKYIVFAHGSFASNNCHCCHKEMETETIKTYMKDKKIPSCQHC-----GYVKPDIVVF 175
 Qy 192 GENLPEQHRANKYDKDEVLLIVIGSSLKVRPALIISSTIHEVPQILINER 244
 Db 176 GGLPVPKPFDMEDDCEDEVAIVAGTSLTFVPPASLPEVANKKCLRLVYNKE 228

RESULT 9
 S59678
 HST2 protein - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein LPA2c; protein YPL015c
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
 C:Accession: S59678
 R:Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; W.
 submitted to the EMBL Data Library, August 1995
 A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
 A:Reference number: S59677
 A:Accession: S59678
 A:Molecule type: DNA
 A:Residues: 1-357 <HAL>
 A:Cross-references: EMBL:U03335; NID:g965076; PIDN:AAB68090.1; PID:g965078; MIPS:YPL015c
 C:Genetics:
 A:Gene: SGD:HST2
 A:Cross-references: SGD:S0005936; MIPS:YPL015c
 A:Map position: 16L

C:Keywords: transmembrane protein
 F:218-234/Domain: transmembrane #status predicted <TMM>

Query Match 32.5%; Score 418; DB 2; Length 357;
 Best Local Similarity 40.2%; Pred. No. 5.3e-30;
 Matches 92; Conservative 39; Mismatches 78; Indels 20; Gaps 5;

Qy 32 KIIVLTGAGVSVSCGIPDRS-RDGIYARLAVDPDLPPQAMFIEYFRKDPPEFFAK 90
 Db 26 KIIFWVGAGISTSCGIPDRSGTGLYHFLA--RLKLPPEAVFVDFQSPPLFFITLA 83
 Qy 91 KEIYPGQFSLCHKFIALSDKEGLLNRYTONIDTLEQVAGI--RILQCHGSFATASC 148
 Db 84 KELYPGNFPRSKFHYLLKLFQDKDVLRVYTONIDTLEQAGVCKDILLIEAHGSFAHCHC 143
 Qy 149 LICKKYKVDCEAVRGDIFNOVVRPCRPCRADEPLAIKPEIYFEGENLPEQHRANKYDKD 208
 Db 144 ICGKRVYPPQVPRSKLAEHPINDFYKC--DYCGELVKPAIVFEGDLPDSFSETWLNSE 201
 Qy 209 -----EVDLIVIGSSLKVRPALIISSTIHEVPQILINER 244
 Db 202 WLREKITTSGKHPQDPLVIVGTSLAVIPFASLPEEIPKRVKRVLCNLE 250

RESULT 10
 A72370
 regulatory protein, SIR2 family - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: A72370
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72001; MUID:99287316
 A:Accession: A72370
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-246 <ARN>
 A:Cross-references: GB:AE001726; GB:AE000512; NID:94980992; PIDN:ABD35575.1; PID:9498
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM0490
 C:Superfamily: conserved hypothetical protein b1120

Query Match 26.9%; Score 346; DB 2; Length 246;
 Best Local Similarity 37.3%; Pred. No. 1e-23;
 Matches 79; Conservative 41; Mismatches 76; Indels 16; Gaps 4;

Qy 20 TIEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSGRTIYARLAVDPDLPPQAMFIEYF 79
 Db 3 MKEFLDLNESRLTYTLTGAGISTGSPGIDPFGPNGIYKYS-----QNVEDIDFF 53
 Qy 80 RQDPPEFFAKEI-YPG-QFQPSLCHKFIALSDKEGLLNRYTONIDTLEQVAGIORTL 137
 Db 54 YHPPEFFYFAKEGIFPMLOAKPNLAHYLALEKGLIEAVIYTONIDTLEHQRASGKVT 113
 Qy 138 QCHGSFATASCLICKKYKVDCEAVRGDIFNOVVRPCRPCRADEPLAIKPEIYFEGENLPE 197
 Db 114 ELHGNVEEYCYCRCKKYVEIVYIKLESDDVPLDCCN-----SLIRNIVYFEGENLPQ 168
 Qy 196 QFHRANKYDKDEVLLIVIGSSLKVRPALIP 229
 Db 169 DALREAIGLSRSASIMIVIGSSLVVYPAALP 200

RESULT 11
 H69827
 conserved hypothetical protein yhdz - Bacillus subtilis
 C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: H69827

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Beren
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chn
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 350, 249-256, 1997

A:Authors: Foulter, D.; Foltz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koltingstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A.; Authors: Schlach, S.; Schoefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowsky, A.; Sero
A.; Authors: Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Walpac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Darchin, A.

A>Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: H69827

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-247 <KUN>

A:Cross-references: GB:Z99109; GB:AL009126; NID:92633260; PIDN:CAB12804.1; PID:92633300

A:Experimental source: strain 168

C:Genetics:

A:Gene: yhdz

C:Superfamily: conserved hypothetical protein b1120

Query Match 25.7%; Score 331.5; DB 2; Length 247;

Best Local Similarity 34.5%; Pred. No. 2e-22;

Matches 81; Conservative 48; Mismatches 83; Indels 23; Gaps 8;

QY 20 IEDAVKLOECKKIIVLTGAGVSCGIDPFRSDGIYARLAVDPDLPQAMFIEYF 79

DB 1 MEFFKSLIHEAQRIYVLTGAGMSTESGIDPFRSAGIWEF-----DASHMEAN-SLDYF 53

QY 80 RNDPRFFFAKRIY----PGORPQLCHKFTLSDEKGLLNNTQNTDLEQVAGIOR 135

DB 54 LSTPLFWEKFEKLEFQMKMGSFEPNEGHLLELEKQKQVDFITQNDLGHKKGSRH 113

QY 136 ILQCHGFATASCLICKKYKDCFAVNGDIFNQVPPRCPCPADPEPL--AIMEPEIVFGE 193

DB 114 VIELHSGISQTACRCAGARYDLR----HLIERVEPTAAGNNGDLCGYLTKDVLFGD 169

QY 194 NPEQFHRAMKYDK-DEVLLIVIGSSLVKVPVALIP--SSIPHEVQQLINREP 245

DB 170 AV---MHFPLVEKRLDQADLLVIGTSLLEVAPARFVPEASLIRPKMKVIINREP 221

RESULT 12

H69263

transcription regulator sir2 family homolog - *Archaeoglobus fulgidus*

C:Species: *Archaeoglobus fulgidus*

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: H69263

R:Kleek, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

Giolek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kalne, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343

A:Accession: H69263

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-253 <KLE>

A:Cross-references: GB:AE001098; GB:AE000782; NID:92689421; PIDN:AAB91115.1; PID:9265053

C:Superfamily: conserved hypothetical protein b1120

Query Match 24.7%; Score 318; DB 2; Length 253;

Best Local Similarity 33.8%; Pred. No. 3.4e-21;

Matches 77; Conservative 38; Mismatches 99; Indels 14; Gaps 4;

QY 20 IEDAVKLOECKKIIVLTGAGVSCGIDPFRSDGIYARLAVDPDLPQAMFIEYF 79

DB 5 IRRAAELIASKRAVAVFTGAGISGIPFREDGIMRY-----DPEEVAISISGF 56

QY 80 RNDPRFFFAKRIYGOQ--QPSLCRKFITLSDEKGLLNNTQNTDLEQVAGIOR 138

DB 57 KRNPRFWEFMSMKKLFAPNPAPAHYALAELERMGIVKAVITQNDIMLHQRGSRVLE 116

QY 139 CHGSFATASCLICKKYKDCFAVNGDIFNQVPPRCPCPADPEPLAIKKPEIVFGENLPEQ 198

DB 117 LHSMSKLDCLDCHETIYDMESEFEDRNKGITPRCKCGS-----YYKPRVLYFGEPLPQR 172

QY 199 FHRAMKYDKDEVLLIVIGSSLVKVPVALIPSSIPHE-VQQLINREP 245

DB 173 TLFEALIEAKHCDAPFVVGSSLVVYPAELPYIAKKAGAKMIVNAEP 220

RESULT 13

S54631

HST3 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein 02651; protein YOR025w

C:Species: *Saccharomyces cerevisiae*

C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999

C:Accession: S54631; S62109; S66891

R:de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54617

A:Accession: S54631

A:Molecule type: DNA

A:Residues: 1-447 <DEH>

A:Cross-references: EMBL:X87331; NID:q1041652; PIDN:CAA60741.1; PID:9829135

R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D

submitted to the EMBL Data Library, October 1995

A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in

A:Reference number: S62108

A:Accession: S62109

A:Molecule type: DNA

A:Residues: 1-447 <BRA>

A:Cross-references: EMBL:U39062; NID:q1055021; PIDN:AAA81034.1; PID:q1055022

A:Experimental source: strain YPH1 (S288C)

R:de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66877

A:Accession: S66891

A:Molecule type: DNA

A:Residues: 1-447 <DEM>

A:Cross-references: EMBL:Z74933; NID:q1420134; PIDN:CAA99215.1; PID:e251955; PID:q142

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:HST3

A:Cross-references: SGD:S0005551; MIPS:YOR025w

A:Map position: 15R

Query Match 24.5%; Score 315; DB 2; Length 447;

Best Local Similarity 32.0%; Pred. No. 1.3e-20;

Matches 91; Conservative 37; Mismatches 96; Indels 60; Gaps 11;

QY 16 DINTIEDAVKLOECKKIIVLTGAGVSCGIDPFRSDGIYARLAVD---FPDLPDPA 72

DB 38 DDEVLRKVTQQLSRSRRIACITGAGISGIPFREDGIMRY-----DPEEVAISISGF 97

QY 73 MPDIEYFRKDP--PEPKFAKRIYGP--QPSLCRKFITLSDEKGLLNNTQNTDLEQVAGIOR 128

DB 98 MPDISLFRDQKISIFAKFERLYSNVQLAKPTAKFTIAHLKDRNKLKLCYQNDIGLE 157

QY 129 QVAGIQ-----RIQCHGSPATASCLIC-----KKYKDCFAVNGD 163

DB 158 ESIGILTSLNKKLPLTSSHWKMLNDVYQLHGDLKTLISCTCGCFQFPMSRYWNRCLARGE 216

QY 164 IFNOVPRPCPCPA-----DEPLAIKPEIVFGENLP--EQFHRAMKYD-- 206
 Db 217 L-----PLCPCECALINKRLNEGRKTLGNSGILRPNIYLGEHNPSCETITGGINLDII 271
 QY 207 KDEVDLLIVGSSSLKVRPVALIPSSIPHEVPQ-----ILLIREP 245
 Db 272 KGNPDFLIMGTSLKVDGKOLVKRLKRIHGRGLITLVNKP 315

RESULT 14

C69459

transcription regulator Sir2 family homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: C69459

R:Kienk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea

A:Reference number: A69250; MUID:98049343

A:Accession: C69459

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-245 <KLE>

A:Molecule type: DNA

A:Cross-references: GB:AE000987; GB:AE000782; NID:92689310; PIDN:AAB89569.1; PID:9264887

C:Superfamily: conserved hypothetical protein b1120

Query Match 22.3%; Score 287.5; DB 2; Length 245;
 Best Local Similarity 31.3%; Pred. No. 1.8e-18;
 Matches 73; Conservative 46; Mismatches 79; Indels 35; Gaps 9;

QY 21 EDANKLLQECKKIIVLTGAGVSVCIGIPDFSRDGIYARLAVDFPDLPPQAMFIEYFR 80
 Db 3 EKLKTTAESKYVALTGAGVASGSIPTFRKDGIMNRYRPE--ELANPQA-----FA 54
 QY 81 KDPAPPFKE----AKEIYPGQPSLCHKFIALSDKEGKLLRNTQNTDLEQVAGIQRI 136
 Db 55 KDPKVKWKVYMRMEKVF--NAQPNKAQAFAELERGLVKCLITQNVDDLHERAGSRNV 112
 QY 137 LQCHGSFATASCLICKKYKVDCEAVRGDIFNQVPRPCPADDEPLAIKPEIVFGENL- 195
 Db 113 IHLGSLRVVCTSCNNSFEVESA-----PKIPPLPKC--DKCGSLRPGVWVEGEMLP 164
 QY 196 PEQFHRAMKYDKDEVLLIVGSSSLKVRPVALLIPSSI-----PHEVP 237
 Db 165 PDVLDRAMR-EVERADVIIVAGTSAVQPAASLPLIVKORGAIIEINPDETP 216

RESULT 15

F71085

hypothetical protein PH0947 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: F71085

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Seki
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
 DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: F71085

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <RAM>

A:Cross-references: GB:AP000004; NID:93236131; PIDN:BAA30044.1; PID:93257361

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by Genbank

C:Genetics:

A:Gene: PH0947

C:Superfamily: conserved hypothetical protein b1120

Query Match 21.9%; Score 282.5; DB 2; Length 249;
 Best Local Similarity 31.6%; Pred. No. 5.2e-18;
 Matches 68; Conservative 52; Mismatches 72; Indels 23; Gaps 9;

QY 22 DAVKLLQECKKIIVLTGAGVSVCIGIPDFSRDGIYARLAVDFPDLPPQAMFIEYFR 81
 Db 4 EVARVLASSKNVIAFTGAGISAESGIPFRKDGIMNRYRPE-----PELATPPEAFER 55
 QY 82 DPR---PEFRF-AKEIYPGQPSLCHKFIALSDKEGKLLRNTQNTDLEQVAGIQRI 137
 Db 56 NPKLWDFYKWKRIKIL--KAKPNPAHAYALELKMGLKAVITQNVDDLHREGATNLI 113
 QY 138 QCHGSFATASCLICKKYKVDCEAVRGDIFNQVPRPCPADDEPLAIKPEIVFGENL 195
 Db 114 ELHGNIFRVCTSCFPRHLKESGRIDELISDDLKPKCKG-----SLRRDVMVFGEP 168
 QY 196 PEQFHRAMKYDKDEVLLIVGSSSLKVRPVALLIP 229
 Db 169 PSKEINERFKLAK-EADVIVVGTSGLVYPAAYIP 202

Search completed: February 16, 2001, 10:45:26
 Job time: 227 sec

Tue Feb 27 15:47:18 2001

us-09-461-580a-4.rpr

Page 7

1
:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:52:25 ; Search time 42.08 Seconds
(without alignments)

188.024 Million cell updates/sec

Title: US-09-461-580A-4

Sequence: 1 VINILSEPPKRRKKRDINTI.....ALIPSIPEVQQLINREP 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551	42.8	670	1	SIR2_KLUOA
2	535.5	41.6	503	1	HST1_YEAST
3	518	40.2	562	1	SIR2_YEAST
4	507.5	39.4	515	1	SIR2_CANAL
5	470	36.5	381	1	SIR2_LEIMA
6	448	32.5	357	1	HST2_YEAST
7	418	24.5	447	1	HST3_YEAST
8	229	17.8	370	1	HST4_YEAST
9	97	7.5	708	1	TRFL_BOVIN
10	91	7.1	708	1	TRFL_BUBAR
11	82.5	6.4	449	1	GATB_METTH
12	82.5	6.4	606	1	ALB1_XENLA
13	81.5	6.3	1102	1	YE20_MENJA
14	79.5	6.2	1010	1	ATN3_CHICK
15	78.5	6.1	1411	1	YMA2_YEAST
16	78	6.1	409	1	MASS_YEAST
17	77.5	6.0	1013	1	ATN3_HUMAN
18	77.5	6.0	1013	1	ATN3_RAT
19	77.5	6.0	2366	1	TOXB_CLODI
20	77	6.0	312	1	MIAA_HELPJ
21	77	6.0	471	1	ATPB_HERAU
22	77	6.0	1165	1	PEX6_PICPA
23	76	5.9	412	1	ALIC_BACSU
24	76	5.9	3859	1	RPOA_LELV
25	75	5.8	703	1	YD94_METTH
26	75	5.8	708	1	TRFL_CAPIH
27	75	5.8	758	1	LEU2_SCHPO
28	74.5	5.8	396	1	ARRS_XENLA
29	74.5	5.8	429	1	AS11_YEAST
30	74.5	5.8	1788	1	TP72_CABEL
31	74	5.7	157	1	RL12_CHIRE
32	74	5.7	297	1	CC2_SCHPO
33	74	5.7	314	1	ARCC_CLOPE

34	74	5.7	444	1	TIG_MYCPN	P75454 mycoplasma
35	74	5.7	546	1	GCVR_HSV7J	P52344 herpes simp
36	74	5.7	607	1	ALB2_XENLA	P14872 xenopus lae
37	73.5	5.7	497	1	RP54_PSEPU	P15591 pseudomonas
38	73.5	5.7	838	1	OS94_MOUSE	P48732 mus musculu
39	73.5	5.7	1274	1	MYPC_HUMAN	Q14896 homo sapien
40	73	5.7	233	1	HXA5_RAT	P52949 rattus norv
41	73	5.7	450	1	YEGD_ECOLI	P36938 escherichia
42	73	5.7	510	1	CBP1_ORISA	P37880 oryza sativ
43	72.5	5.6	348	1	CHLI_EUGLEA	P31205 euglena gra
44	72.5	5.6	620	1	ORC2_YEAST	P32833 saccharomyc
45	72.5	5.6	863	1	MCM4_HUMAN	P33991 homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	670 AA.
SIR2_KLUOA				
ID	SIR2_KLUOA			
AC	P33294			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DE	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).			
GN	SIR2.			
OS	Kluyveromyces fragilis (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;			
OC	Saccharomycetaceae; Kluyveromyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CBS 2359/1152;			
RX	MEDLINE-94277055; PubMed-8007956;			
RA	Chen X.-D., Clark-Walker D.G.;			
RT	"SIR2 mutants of Kluyveromyces fragilis are hypersensitive to DNA-			
RT	targeting drugs.";			
RL	Mol. Cell. Biol. 14:4501-4508(1994)			
CC	-1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM			
CC	MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS			
CC	INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI			
CC	SILENCING (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X74569; CAAS2661.1; -			
DR	PIR: S36616; S36616.			
KW	Transcription regulation; Repressor; DNA-binding; Zinc-finger;			
KW	Nuclear protein.			
FT	DOMAIN 44			
FT	DOMAIN 310			
FT	ZN_FING 428			
FT	DOMAIN 650			
FT	POLY-SER.			
SQ	SEQUENCE 670 AA; 74227 MW; 00A063C743308B75 CRC64;			

Query Match 42.8%; Score 551; DB 1; Length 670;

Best Local Similarity 44.6%; Pred. No. 2.2e-42;

Matches 119; Conservative 41; Mismatches 73; Indels 34; Gaps 5;

QY 11 RKKRNDITDAVKLLQCKKIVIVGAGVSCGIPDFSRGCIYARLAVDPDPDP 70
DB 291 RIRLTNHTTIDDFAKLTAKRIIVLGAGISTGIPDFSRSGFYSKLD-DL-GLNDP 348
QY 71 QAMPDIEFRKDPPEFFAKIYPGOPSLCKKFTALSDKEKLLRNYQNTDTLEOV 130

DB 349 ODVLEFTEPDSVFYNAHVPENNYSPLSHFIKMIDKRLNRYTONIDNLESY 408
 QY 131 AGI--QRILQCHGFATASCLCKKYKVDCEAVRGDIFNOVRCPRC-----PAD 178
 DB 409 AGVEPEKMWQCHGSPATACVCHWKIGERIFPNRNQLTICPCYCKRLEFRTKTD 468
 QY 179 EPLA-----IMKPEIVFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVR 218
 DB 469 EBLAGDEDDMDHGRSVKPSFGVLKPDITFGEALPSKFRHLRREDVLCDDLLICIGT 528
 QY 219 SLKVRVALIPSSIPHEVQOLLINREP 245
 DB 529 SLKVRVSEIVNMIPAHVPOVLINKDP 555

RESULT 2
 HST1_YEAST
 ID HST1_YEAST STANDARD: PRT: 503 AA.
 AC P53685;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST1 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 1).
 GN HST1 OR YOL068C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPH1;
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 Boeke J.D.;
 RT "Type SIR2 gene family, conserved from bacteria to humans, functions
 in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GRY 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Derbyshire M.K., Weinstein K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97321807; PubMed-9178509;
 RA Tzeremla M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast
 chromosome XV reveals eight known genes and ten new open reading
 frames including homologues of ABC transporters, inositol
 phosphatases and human expressed sequence tags.";
 RL Yeast 13:583-589(1997).
 CC -1- FUNCTION: INVOLVED IN TELOMERIC SILENCING, AS OVEREXPRESSION
 RESTORES SILENCING AT HMR IN SIR2 MUTANTS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: U39041; AAA81033.1;
 DR EMBL: L47120; AAB38430.1;
 DR EMBL: Z74810; CAA99078.1;
 DR SGD: S0005429; HST1.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT DOMAIN 200 439 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 318 345 C4-TYPE (POTENTIAL).

SQ SEQUENCE 503 AA; 57702 MW; 4CDF2799E4135ABB CRC64;
 Query Match 41.6%; Score 535.5; DB 1; Length 503;
 Best Local Similarity 42.9%; Pred. No. 4e-41;
 Matches 112; Conservative 43; Mismatches 77; Indels 29; Gaps 3;

QY 11 RKRKDINTEEAVALLOECKKIYVTGAGVSQSGIPDFRSRDGIYARLAVFPDLPDP 70
 DB 181 RLRLNENFTIDFTALRNARKIIVLTGAGVSTSLGIPFRSSEGFYSK--IRHGLDEP 238
 QY 71 QAMFDEYFRKDPPEPFKAKETIYEQFQPSLCHKFIASLDEKGLNRYTONIDPLEOV 130
 DB 239 ODVFLDIFLOPDSVFYNAHVPENNYSPLSHFIKMIDKRLNRYTONIDNLESY 298
 QY 131 AGI--QRILQCHGFATASCLCKKYKVDCEAVRGDIFNOVRCPRC----- 175
 DB 299 AGIDPKLVQCHGSPATACVCHWKIPEKIFENIRNLELPLCPAGYCKRRKQYFPMMSG 358
 QY 176 -----PADEPLAIMKPEIVFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVR 223
 DB 359 NNTVQTNINENSPILKSYVLKPDMTFGEALPSRFRKIRRDILBCDLLICIGSLKVA 418
 QY 224 PVALIPSSIPHEVQOLLINREP 244
 DB 419 PVSEIVNMVPSHVPOVLINKDP 439

RESULT 3
 SIR2_YEAST
 ID SIR2_YEAST STANDARD: PRT: 562 AA.
 AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 GN SIR2 OR MAR1 OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85126876; PubMed-6098447;
 RA Shore D., Squire M., Naamlych K.A.;
 RT "Characterization of two genes required for the position-effect
 control of yeast mating-type genes.";
 RL EMBO J. 3:2817-2823(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Paulin L., Saren A.M., Laamanen P.;
 RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 SILENCING. SUPPRESSES RDNA RECOMBINATION AND PROMOTES HISTONE
 DEACETYLATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: X01419; CAA25667.1;
 DR EMBL: Z71781; CAA96447.1;
 DR EMBL: Z74090; CAA98600.1;
 DR PIR: S05891; RGSYS2.
 DR SGD: S0002200; SIR2.

KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 254 498 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 372 399 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 562 AA; 63261 MW; 526693753654586 CRC64;

Query Match 40.2%; Score 518; DB 1; Length 562;
 Best Local Similarity 39.1%; Pred. No. 1.8e-39;
 Matches 117; Conservative 44; Mismatches 80; Indels 58; Gaps 7;

QY 1 VINLSEPPKRRK-----KDIN-----TIEDAVKLLQCKKIIVLTG 38
 DB 205 VTNV--EDPLAKQTVARLIDQRAINKVLCRLRLSNFTIDHFIQKLTARKILVLTG 262
 QY 39 AGVSVCGLIDFERSRDCITVARLAVDPDLPDPOAMFIEFRKPPREFFAKITYGQF 98
 DB 263 AGVSTSLGIDFERSSEGEYSK--IKHLGLDDPDQVFNFMHDPVFNNTANNVLPPEK 320
 QY 99 QPSLCHKFIALSDKEGKLLNNTQNTIDTLEOVAGI--QRILQCHSPATASCLICKYKD 156
 DB 321 IYBLSHFIMLOMKGKLLNNTQNTIDTLEOVAGI--QRILQCHSPATATCVTCHMNL 380
 QY 157 CEAVRGDIFNQVPRCPRC-----PADP-----LAIMKP 186
 DB 381 GERIFNKRIRLLELPCLPCYCKRRREYPEGNNKVGVAASGSGSEKRPYILNSGYLKP 440
 QY 187 EIVPEGNELEQFRAMKRYKDEVDLLIVGSSLKRPVALIPSSIFHEVQIILNREP 245
 DB 441 DIFEGALPNKFKHSIREDLLECDDLICIGTSLKAVPASEIVMVPSPQVILNRP 499

RESULT 4
 SIR2_CANAL STANDARD; PRT; 515 AA.
 ID SIR2_CANAL
 AC 059923;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE REGULATORY PROTEIN SIR2.
 GN SIR2.
 OS Candida albicans (yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC anamorphic Saccharomycetales; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martín J., Johnson A.D.;
 RT "SIR2 gene from Candida albicans."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF045774; AAC09304.1;
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 FT Nuclear protein.
 FT DOMAIN 234 453 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 352 379 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 515 AA; 57732 MW; 2E4DB259EFA9251F CRC64;

Query Match 39.4%; Score 507.5; DB 1; Length 515;
 Best Local Similarity 45.9%; Pred. No. 1.4e-38;
 Matches 107; Conservative 41; Mismatches 76; Indels 9; Gaps 3;

QY 20 IEDAVKLLQCKKIIVLTGAGVSCGIPDFERSHGIYARLAVDPDLPPOAMFIEYF 79
 DB 224 LPDLISLRKKIMVYTGAGISLSDIPDRSPKGLYNLSK--LNLSDPKYFDLOT 281
 QY 80 RKDPRPFKFAKEIYPOFQPSLCHKFIALSDKEGKLLNNTQNTIDTLEOVAGI--QRIL 137
 DB 282 MREGRLFYTTANHLVLPDQKFSILHAFKLKLDQKHKLRYNTQNTIDTLEOVAGI--QRIL 341
 QY 138 QCHSEFATSLCKIKYVNDQCAVVRGDIENQVPRCPRC-----PADPRLAIMEKEIYFG 192
 DB 342 QCHSEFATSLCKIKYVNDQCAVVRGDIENQVPRCPRC-----PADPRLAIMEKEIYFG 401
 QY 193 ENLPEQFRAMKYKDEVDLLIVGSSLKRPVALIPSSIFHEVQIILNREP 245
 DB 402 EDLPERRHTLMDKLDQDIDFLVIGTSLKVPASIIERYVYKPKILINRP 454

RESULT 5
 SIR2_LEIMA STANDARD; PRT; 381 AA.
 ID SIR2_LEIMA
 AC 025337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RP).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96186914; Pubmed-8635734;
 RA Yehiouli B., Tahbi A., Ouassil A.;
 RT "A Leishmania major protein with extensive homology to silent
 RT information regulator 2 of Saccharomyces cerevisiae."
 RL Gene 169:115-118(1996).
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: LA0331; AAB06804.1;
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 FT Nuclear protein.
 FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 148 151 POLY-ALA.
 FT ZN_FING 152 179 C4-TYPE (POTENTIAL).
 FT DOMAIN 271 278 POLY-SER.
 FT DOMAIN 305 310 POLY-SER.
 SQ SEQUENCE 381 AA; 41958 MW; 5A311630A5D2365 CRC64;

Query Match 36.5%; Score 470; DB 1; Length 381;
 Best Local Similarity 41.3%; Pred. No. 2.5e-35;
 Matches 102; Conservative 45; Mismatches 76; Indels 24; Gaps 7;

QY 5 LSEPPKRRKRDINTIEDAVKLLQE--CKRIIVLTGAGVSCGIPDFERSD-GIYARLA 61
 DB 14 LGEP-----TVGLAKRYIREKQVRITLVAGAGAVAAAGIIPDFRSRDTGIYAKLG 63

```

QY 62 VDEPDLPOQAMFIEYFRKDRPPEFKAKE--IYGOFOPSLCHKFTALSDKEGKLLRN 119
DB 64 K--YMLDPTDAFSLTLRKEPEFYIARLMLNPGHOPFAVHFRLLQDEGRLLK 121
QY 120 YTONIDTLEQVAGT--QRILOCHGSFATASCLICRYKDCENVRQDINQVPRCPRA 177
DB 122 CTQNDIGLEKAGVSPELLVEAHGSAACIECHTPESEIENYLEANSQVRSCTGC- 180
QY 178 DEFLAIKPEIYFEGENLPEQFHRAMKYDKDEVDLLIYGSLSKVRPALPISIPHEVP 237
DB 181 ----GIVAPNVYFEGENLPDAFFDALHHDAPLAEIYIIGISMVHPFALLPCVVPKSI 236
QY 238 QILINRE 244
DB 237 RYLMNRE 243

RESULT 6
HST2_YEAST STANDARD: PRT: 357 AA.
AC P53686:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST2 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 2).
GN HST2 OR YPL015C OR LPAC2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1.
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
  Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
  in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Hall J., Ahmed A., Bussey H., Fortin N., Friesen J.D., Storms R.K.,
  Vo D.H., Wang Y., Winnett E.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: COULD FUNCTION IN TELOMERIC SILENCING, CELL CYCLE
  PROGRESSION AND CHROMOSOME STABILITY.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39063; AAA81035.1; -
DR EMBL: U33335; AAB68090.1; -
DR SCD: S0005936; HST2.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
  Nuclear protein.
FT ZN-FING 143 173 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 143 173 C4-TYPE (POTENTIAL).
SQ SEQUENCE 357 AA; 39979 MW; ED281E5B8241A4D0 CAC64;

```

```

Query Match 32.5%; Score 418; DB 1; Length 357;
Best Local Similarity 40.2%; Pred. No. 1.2e-30;
Matches 92; Conservative 39; Mismatches 78; Indels 20; Gaps 5;

```

```

QY 32 KIIVTGTAGVSVSCGIPDRSRDGIYARLAVDPDLPOQAMFIEYFRKDRPPEFKFA 90
DB 26 KVIEMVAGISTSCGIPDRSPGTGLYHNLA--RLKLPYPAEAVFDVDFOSDPLPFYTLA 83

```

```

QY 91 KEIYGOFOPSLCHKFTALSDKEGKLLRNQTONIDTLEQVAGT--RILOCHGSFATASC 148
DB 84 KELYGNFRSPSFHYLLKLFQDKDVKLRVYTONIDTLERQACVKNDDLIIEAGSFAHCHC 143
QY 149 LICKRYKDCENVRQDINQVPRCPRADEPLAIKPEIYFEGENLPEQFHRAMKYDKD 208
DB 144 ICGCKVPPQVFKSKLAEHPIDKFPVC--DVCGELFAPKPAIYFEGEDLPDSFSETWLNDS 201
QY 209 -----EVDLLIYGSLSKVRPALPISIPHEVPQILINRE 244
DB 202 WLREKITTSGRHPQPLIIVGTSLAVYPFASLPPEIIRKRYVLCLNE 250

RESULT 7
HST3_YEAST STANDARD: PRT: 447 AA.
AC P53687:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST3 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
GN HST3 OR YOR025W OR OR26.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1.
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
  Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
  in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RA de Haan M., Maarse A.C., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
  PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
  AND GENOMIC MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39062; AAA81034.1; -
DR EMBL: X87331; CAA60741.1; -
DR EMBL: Z74933; CAA99215.1; -
DR SCD: S0005551; HST3.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
  Nuclear protein.
FT ZN-FING 50 314 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 195 223 C4-TYPE (POTENTIAL).
SQ SEQUENCE 447 AA; 50524 MW; 59CDS5FBD712B7005 CAC64;

```

```

Query Match 24.5%; Score 315; DB 1; Length 447;
Best Local Similarity 32.0%; Pred. No. 3.3e-21;
Matches 91; Conservative 37; Mismatches 96; Indels 60; Gaps 11;

```

```

QY 16 DINTEDAVKLLQECKKIIVTGTAGVSVSCGIPDRSRDGIYARLAVP---FPDLPPQQA 72
DB 38 DDEIVLRVTKOLSRSRRIACLTGAGISCNAGIPDRSSDGLYLVKKDCSOYWSIKSGRE 97

```



```

RT multiple molecular weight forms."
RL Comp. Biochem. Physiol. 93B:929-934(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=98063267; PubMed=9398529;
RA Moore S.A., Anderson B.F., Groom C.R., Haridas M., Baker E.N.;
RT "Three-dimensional structure of ferric bovine lactoferrin at 2.8-A
  resolution."
RL J. Mol. Biol. 274:222-236(1997).
RN [7]
RP STRUCTURE BY NMR OF 36-60.
RX MEDLINE=98190007; PubMed=9521752;
RA Hwang P.M., Zhou N., Shan X., Arrowsmith C.H., Vogel H.J.;
RT "Three-dimensional solution structure of lactoferrin B, an
  antimicrobial peptide derived from bovine lactoferrin."
RL Biochemistry 37:4288-4298(1998).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- FUNCTION: LACTOFERRICIN B IS AN ANTIMICROBIAL PEPTIDE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; X57084; CAA40366.1; -.
DR EMBL; M63502; AAA30617.1; -.
DR EMBL; L08604; AAA30609.1; -.
DR EMBL; L19993; AAA21722.1; -.
DR EMBL; L19982; AAA21722.1; JOINED.
DR EMBL; L19983; AAA21722.1; JOINED.
DR EMBL; L19984; AAA21722.1; JOINED.
DR EMBL; L19985; AAA21722.1; JOINED.
DR EMBL; L19986; AAA21722.1; JOINED.
DR EMBL; L19988; AAA21722.1; JOINED.
DR EMBL; L19989; AAA21722.1; JOINED.
DR EMBL; L19990; AAA21722.1; JOINED.
DR EMBL; L19991; AAA21722.1; JOINED.
DR EMBL; L19992; AAA21722.1; JOINED.
DR PIR; S14674; TFBOL.
DR PIR; JTO595; JTO595.
DR PDB; 1BLF; 03-DEC-97.
DR PDB; 1LFC; 18-NOV-98.
DR INTERPRO; IPR001156; -.
DR PFAM; PFO0405; transferf; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Iron transport; Glycoprotein; Metal-binding; Repeat; Signal;
KW Antibiotic; 3d-structure.
FT SIGNAL 1 19
FT CHAIN 20 708 LACTOTRANSFERRIN.
FT PEPTIDE 36 60 LACTOFERRICIN B.
FT REPEAT 20 363 1.
FT REPEAT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.

```

```

FT DISULFID      444    666     BY SIMILARITY .
FT DISULFID      476    551     BY SIMILARITY .
FT DISULFID      500    694     BY SIMILARITY .
FT DISULFID      510    524     BY SIMILARITY .
FT DISULFID      521    534     BY SIMILARITY .
FT DISULFID      592    606     BY SIMILARITY .
FT DISULFID      644    649     BY SIMILARITY .
FT METAL         79      79      IRON 1 .
FT METAL        111     111     IRON 1 .
FT METAL        211     211     IRON 1 .
FT METAL        272     272     IRON 1 .
FT METAL        414     414     IRON 2 .
FT METAL        452     452     IRON 2 .
FT METAL        545     545     IRON 2 .
FT METAL        614     614     IRON 2 .
FT BINDING       140     140     ANION (POTENTIAL) .
FT BINDING       482     482     ANION (POTENTIAL) .
FT CARBOHYD      252     252     N-LINKED (GLCNAC . . ) (POTENTIAL) .
FT CARBOHYD      387     387     N-LINKED (GLCNAC . . )
FT CARBOHYD      495     495     N-LINKED (GLCNAC . . )
FT CARBOHYD      564     564     N-LINKED (GLCNAC . . )
FT CONFLICT      63      63      E -> A (IN REF. 4) .
FT CONFLICT      66      67      RA -> PG (IN REF. 2) .
FT CONFLICT      145     145     I -> V (IN REF. 1 AND 4) .
FT CONFLICT      164     165     LO -> PR (IN REF. 1) .
FT CONFLICT      264     264     C -> Y (IN REF. 4) .
FT CONFLICT      273     273     A -> P (IN REF. 4) .
FT CONFLICT      281     281     G -> A (IN REF. 4) .
FT CONFLICT      291     291     S -> R (IN REF. 4) .
FT CONFLICT      297     297     F -> S (IN REF. 2) .
FT CONFLICT      340     340     G -> A (IN REF. 1) .
FT CONFLICT      418     418     I -> V (IN REF. 4) .
FT CONFLICT      439     439     H -> Y (IN REF. 1) .
FT CONFLICT      459     459     K -> R (IN REF. 4) .
FT CONFLICT      514     514     A -> R (IN REF. 1) .
SQ SEQUENCE      708 AA;   78056 MM; CCFD/FCLSD6B93FC CRC64;

Query Match          7.5%; Score 97; DB 1; Length 708;
Best Local Similarity 23.6%; Pred.No. 0.33;
Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

OY 81 KDPRFFFAKFIYGCQOPSLCHHFIALSK-----BGK----- 115
DB 88 RDPLYRLPAVAEITGYTKSPQTHTYAAVAIVVKKGSNFOLDOLGKRKSCHTGLGRSAGWIIP 147
OY 116 --LLENTNYONTDLTEBOVGVIQRILLQHSGSFATASCLIC---KYKVDECAVVRGDIFNOVV 169
DB 148 MGILRPYLTSWISTESLPLOG-----AVAKFSBASCVPCIDRAIYNLCOLCKGEENOCA 201
OY 170 PCRPCPADPEP-----LAIMKEPIVEFGENTLBPQPHRAMKYRDKDEVLD 212
DB 202 -----CSSREPYFGYGSAFKQLDGADGVAFVKETTTF--EWLP-E-----KADRDOVEL 248
OY 213 LIVIGS-----SLKVPVALIPS 230
DB 249 LCLNNSRAPVDAAFEKHCLAOWPS 271

RESULT 10
TRFL_BUBAR
ID TRFL_BUBAR STANDARD; PRT; 708 AA.
AC 077698;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LACTOFTRANSFERRIN PRECURSOR (LACTOFERRIN).
GN LFP.
OS Bubalus arnee bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mamalia; Eutheria; Cetartiodactyla; Rumiantia; Pecora; Bovidae;
NC Bovidae; Bovinae; Bubalus.
[1]

```


RP SEQUENCE FROM N.A.
RA Paramasivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
RT Singh T.P.;
"CDNA sequence of Buffalo lactoferrin.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA Karthikeyan S., Paramasivam M., Yadav S., Srinivasan A., Singh T.P.;
RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
RT grown at 303 K shows different orientations of the N and C lobes.";
RL Acta Crystallogr. D 55:1805-1813(1999).
CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ005203; CA06441.1; -
CC PDB: 1CE2; 19-MAR-99.
DR INTERPRO: IPRO01156; -
DR PFAM: PF00405; transferrin; 2.
DR PRINTS: PR00422; TRANSFERRIN.
DR PROSITE: PS00205; TRANSFERRIN_1; 2.
DR PROSITE: PS00206; TRANSFERRIN_2; 2.
DR PROSITE: PS00207; TRANSFERRIN_3; 2.
KW Iron transport; Glycoprotein; Metal-binding; Repeat; Signal;
KM 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 708 LACTOTRANSFERRIN.
FT DISULFID 28 64
FT DISULFID 38 55
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 179 202
FT DISULFID 189 200
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 534
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79 IRON 1.
FT METAL 111 111 IRON 1.
FT METAL 211 211 IRON 1.
FT METAL 272 272 IRON 1.
FT METAL 414 414 IRON 2.
FT METAL 452 452 IRON 2.
FT METAL 545 545 IRON 2.
FT METAL 614 614 IRON 2.
FT BINDING 140 140 ANION (POTENTIAL).
FT BINDING 482 482 ANION (POTENTIAL).
FT CARBOHYD 252 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 708 AA: 77729 MW: 082600AAB2F9ACD CRC64;

Query Match 7.1%; Score 91; DB 1; Length 708;
Best Local Similarity 24.9%; Pred. No. 1.1;
Matches 49; Conservative 25; Mismatches 61; Indels 62; Gaps 10;
QY 82 DPREFFKAEIYPGQPSLCHKETALSDK-----EGK-----115
DB 89 DPKRLRVAAEIVGTGKSPQTHYVAVAVKGSNFQDLOGRNSCHTGLGRSAGNIPM 148
QY 116 -LNRNTQNDITLQVAGVIRILOCHSFPATSLIC-----KKVDCFAVRGDFINQV- 169
DB 149 GILRPYLSTWSELEPFG-----AVAKFPSASCVPDRQAVPNLCQLCKGEGENCAC 202
QY 170 -PRCP-----RCPAD--EPLAIMPELVFEGENPEQFHRAMKXDKDQVLDLIYGS 218
DB 203 SPREPTFGYSGARCKLDGAGDAFVETVYF--ENLPE-----KADROYELLCLNNT 254
QY 219 -----SLKVRPVALLIPS 230
DB 255 RAPVDARKECHLAQVPS 271
PRT: 449 AA.
RESULT 11
GATB_METTH STANDARD; PRT: 449 AA.
ID 027341;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE PROBABLE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-)
DE (GLU-ADT SUBUNIT B).
GN GATB OR MTH1280.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H:
RX MEDLINE=98037514; PubMed=9371463; Delonghery C., Lee H.-M., Dubois J.,
RA Smith D.R., Doucette-Stamm L.A., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Aldredge T., Bashirzadeh R., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nollig J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSMUTATION OF MISCIATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE - ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE
CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000893; AAB85762.1; -
DR INTERPRO: IPRO01773; -
DR PFAM: PF01162; PET112; 1.
DR PROSITE: PS01234; GATB; 1.
KW Protein biosynthesis; Ligase.
SQ SEQUENCE 449 AA: 51131 MW: 722F81F2F75BF51 CRC64;

Query Match 6.4%, Score 82.5; DB 1; Length 449;
 Best Local Similarity 26.7%, Pred. No. 3.9;
 Matches 39; Conservative 22; Mismatches 58; Indels 27; Gaps 6;

QY 2 INISEPPKRRKRDINTEDAVKLLQ-ECKKIIVTGVAGSVSCGIDPFRSDGIYARL 60
 DB 177 VNISLEGGRKRVKIKNINSIKGAVKALKEFMIRQKMLKRGVEVKQETRAFLSQMITVSM 236
 QY 61 -----AVDPDPDPPQAMFIDIEFRKDRPFKFAKEIYPGFQ--PSLCH-----KFI-- 107
 DB 237 RLKEAEEDRYRYPDDL-----PPKFTDEVVEGLREIRIPDAPHLKVRMEE 284
 QY 108 -ALSDKGLKRNVTQNTIDTLEOVAG 132
 DB 285 YGLREDARVLTSELDADAEVAVG 310

RESULT 12
 ALBL_XENLA STANDARD; PRT; 606 AA.
 AC P08759:
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE 68 KDA SERUM ALBUMIN PRECURSOR.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 CC Xenopodinae; Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89313788; PubMed=2747653;
 RA Moskatis J.E., Sargent T.D., Smith L.H. Jr., Pastori R.L.,
 RA Schoenberg D.R.;
 RA "Xenopus laevis serum albumin: sequence of the complementary
 RT deoxyribonucleic acids encoding the 68- and 74-kilodalton peptides
 RT and the regulation of albumin gene expression by thyroid hormone
 RT during development.";
 RT Mol. Endocrinol. 3:464-473(1989).
 RN [2]
 RP SEQUENCE OF 1-48 FROM N.A.
 RX MEDLINE=88172470; PubMed=2451026;
 RA Schorpp M., Doebebling U., Wagner U., Ryffel G.U.;
 RT "5'-flanking and 5'-proximal exon regions of the two Xenopus albumin
 RT genes. Deletion analysis of constitutive promoter function.";
 RT J. Mol. Biol. 199:83-93(1988).
 RN [3]
 RP SEQUENCE OF 458-556 FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=85126974; PubMed=3971963;
 RA Wolffe A.P., Glover J.F., Martin S.C., Tenniswood M.P.R.;
 RA Williams J.L., Tata J.R.;
 RA "Deinduction of transcription of Xenopus 74-kDa albumin genes and
 RT destabilization of mRNA by estrogen in vivo and in hepatocyte
 RT cultures.";
 RT Eur. J. Biochem. 146:489-496(1985).
 CC -1- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD
 CC BINDING CAPACITY FOR WATER, CA++, NA+, K+, FATTY ACIDS, HORMONES,
 CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
 CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE ALB/AFIP/VDB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M18350; AAA49636.1; -
 DR PIR: A41682; ABXL68.
 DR PIR: A05288; A05288.
 DR HSSP: P02768; 1A06.
 DR INTERPRO: IPR000264; -
 DR PFAM: PF00273; transport_prot. 3.
 DR PROSITE: PS00212; ALBUMIN; 3.
 KW Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;
 FT SIGNAL. 1 18
 FT PROPEP 19 24
 FT CHAIN 25 606
 FT REPEAT 29 212
 FT REPEAT 218 402
 FT REPEAT 408 600
 FT METAL 30 30
 FT DISULFD 80 89
 FT DISULFD 102 118
 FT DISULFD 117 128
 FT DISULFD 148 193
 FT DISULFD 192 201
 FT DISULFD 224 270
 FT DISULFD 269 277
 FT DISULFD 289 303
 FT DISULFD 302 313
 FT DISULFD 340 383
 FT DISULFD 382 391
 FT DISULFD 414 460
 FT DISULFD 459 470
 FT DISULFD 483 499
 FT DISULFD 498 509
 FT DISULFD 536 581
 FT DISULFD 580 589
 SQ SEQUENCE 606 AA; F8E72AA952C3A3E8 CRC64;

Query Match 6.4%, Score 82.5; DB 1; Length 606;
 Best Local Similarity 22.8%, Pred. No. 5.6;
 Matches 50; Conservative 35; Mismatches 81; Indels 53; Gaps 11;

QY 45 CGIDPFRSDGIYARLAVDFPDLPPQAMFIDIEFRKDRPFKFAK-----EIVPGQFQ 99
 DB 303 CNLP---LLERTYICVLENDVDP-AELSKPTEFTEDPHVCEKAEKSFLEISPMQSQ 358
 QY 100 --PSLCHKFIALSDKEGLLN-----TYQNTIDL-----EQVAGIQR---TLQ 138
 DB 359 ETPELSEQFLLQSAKEYESLNLKCCFSDNPPECKYKGDADRFMNEAKERFAVILKQNCDTILH 418
 QY 139 CGSEFAFASCLICKYKVDCEAVRGDIFNQVPRCPADPEPLAIMKPEIVFGEN--L 195
 DB 419 EHGELFENELLIR-----TKMKQV-SDELILGIANQMDIGSHCAV 462
 QY 196 PEQFHRAMKYKDEYDLIVIGSSLSIKRVPVALLPSSIFH 234
 DB 463 PEN---ORMPCAEGLDTILIGKMCEROKTFFINNVAAH 497

RESULT 13
 YE20_METYA STANDARD; PRT; 1102 AA.
 AC Q58815:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MJ1420.
 GN MJ1420.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

```

RX MEDLINE=96377999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sult G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerevase A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Georgering N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Urtreger T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RA Science 273:1058-1073(1996).
RL -1- PM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: IN THE C-TERMINAL, TO GLUCOSAMINE--FRUCTOSE-6-
CC PHOSPHATE AMINOTRANSFERASES (GFAF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67582; AAB99430.1; -.
CC HSSP; P17169; IMOR.
CC MEROPS; C44.971; -.
CC TIGR; MJ1420; -.
CC INTERPRO; IPR000583; -.
CC INTERPRO; IPR001347; -.
CC INTERPRO; IPR001387; -.
CC INTERPRO; IPR002203; -.
CC PFAM; PF00310; GATASE_2; 2.
CC PFAM; PF01381; HTH_3; 1.
CC PFAM; PF01380; STS; 2.
CC DR PROSITE; PS00443; GATASE_TYPE_II; 1.
CC DR PROSITE; PS00881; PROTEIN_SPLICING; FALSE_NEG.
CC KW Hypothetical protein; Protein splicing.
CC CHAIN 1 74
CC (HYPOTHETICAL PROTEIN MJ1420, 1ST PART
CC (POTENTIAL).
CC CHAIN 75 573
CC MJA G6F INTEIN (POTENTIAL).
CC FT CHAIN 574 1102
CC HYPOTHETICAL PROTEIN MJ1420, 2ND PART
CC (POTENTIAL).
CC SO SEQUENCE 1102 AA; 125907 MW; 9C1EFA8D890CBA1 CRC64;

Query Match 6.3%, Score 81.5; DB 1; Length 1102;
Best Local Similarity 21.4%; Pred. No. 14;
Matches 44; Conservative 29; Mismatches 66; Indels 67; Gaps 9;

OY 11 RKKRDINTIEDAVKLLDECKKIIVL---TGAGSVS-----CGI-----PDRSRDG 55
OY || : : || : : || : : || : : || : : || : : || : : || : : ||
OY 202 RKVREKICLITRKDKVCEKLGKVEIKYIVKTIETGLSEIEEERLKKLCSLGINFEELIYDN 261
OY :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 262 LHYTPVAVFPKPTTPPELMQIIGIYGD-----GHF-PS--NRMRLLDEKKE 305
OY :: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 116 LIRNTQNIIDLTEOVAGIQIRLQCHGSFATASCLICKRYKVDCEAVRGDIFNQVPRCPC 175
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY Db 306 VLEENYQLFKTVENLEG--NIIKKGGNYIILE-INSKYLID----- 343
OY 176 PADEPLAIKPEIVFGENLPEQPIHR 201
OY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY Db 344 -----WFRENIPELFNK 355

RESULT 14
AFN3_CHICK
ID ATN3_CHICK STANDARD: PRT: 1010 AA.
AC P24798;

```

```

DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA-3 CHAIN (EC 3.6.1.37)
DE (SODIUM PUMP) (NA+/K+ ATPASE) (ALPHA(III)).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
RN Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91023019; PubMed=2171348;
RA Takeyasu K., Lemas V., Fambrough D.M.;
RT "Stability of Na(+)-K(+)-ATPase alpha-subunit isoforms in evolution.";
RL Am. J. Physiol. 259:C619-C630(1990).
CC -1- FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME,
CC WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF
CC NA & K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE
CC ELECTROCHEMICAL GRADIENT OF NA & K, PROVIDING THE ENERGY FOR
CC ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
CC -1- CATALYTIC ACTIVITY: ATP + H(2O) = ADP + ORTHOPHOSPHATE.
CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
CC AND GAMMA.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL, M59960; AAA48982.1; -.
DR PIR, B37227; B37227.
DR INTERPRO: IPR000661; -.
DR DR INTERPRO: IPR001757; -.
DR PFAM: PF00122; E1-E2_ATPase; 1.
DR PFAM: PF00689; Na_K_ATPase_C; 1.
DR PFAM: PF00690; Na_K_ATPase_N; 1.
DR PRINTS: PR00119; CATAPASE.
DR PRINTS: PR00121; NAKATPASE.
DR PROSITE: PS00154; ATPASE_E1_E2_1.
KM Hydrolase: Sodium/potassium transport; Transmembrane;
KM Multigene family; Phosphorylation; ATP-binding.
KW TRANSMEM 83 BY SIMILARITY.
FT TRANSMEM 117 136 BY SIMILARITY.
FT TRANSMEM 278 300 BY SIMILARITY.
FT TRANSMEM 307 335 BY SIMILARITY.
FT TRANSMEM 774 797 BY SIMILARITY.
FT TRANSMEM 836 861 BY SIMILARITY.
FT TRANSMEM 903 923 BY SIMILARITY.
FT TRANSMEM 940 965 BY SIMILARITY.
FT MOD_RES 363 363 PHOSPHORYLATION (PROBABLE).
FT BINDING 495 495 ATP (BY SIMILARITY).
SQ SEQUENCE 1010 AA; 111284 MW; 71526BC23633BFA6 CRC64;

Query Match          6.2%; Score 79.5; DB 1; Length 1010;
Best Local Similarity 22.2%; Pred. No. 19;
Matches    58; Conservative   35; Mismatches 101; Indels   67; Gaps      15;

QY 3 NILSEPPRRKKRKDDINTEDAVKLLQECKKI-----VLTGAGVSVCIGIPFR 51
DB 65 NALTTPPP-----TTPEWVKF---CROLFGFSILLMIGAILLCFLAYGIQAQTEDEP 112
QY 52 SHDGIIYARLAUDVPPLDPQANFDLEIYRKRKPREFFKAKETIYPQGFPSLCIKHTIALSD 111
DB 113 SNDNLY--LGIVLAVALVIITGCFSSYYOEAKSKIMESF-KNNVPO-----ALVT 159
QY 112 KEGLKLRYNTQNI---DTLEQVAGIQ-----NILQCHGSFATASCCLICKKYKDCAVBGD 163

```

```
Db 160 REGEMQLNAEEVVDLVEYKGGDRPADLIIISAHG-----C-KVDNSSLTGE 208
OY 164 IFNOVVPKPCRPADPELAIKKPEIVFEGENLPEQFHRAKX---DKXDEVDLLIYGSSL 220
Db - 209 --SEPTSRSPDCTHDNPLETR--NITFFSTNCVECTARGVATADRTVMKRIATLASGL 264
OY 221 KV--RPVALIPSSIPHEVPOI 239
Db 265 EVGKPTPIAV---EIEHFQOLI 282
```

RESULT 15

```
YMA2_YEAST STANDARD; PRT; 1411 AA.
ID YMA2_YEAST STANDARD; PRT; 1411 AA.
AC 003214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 162.7 KDA PROTEIN IN SIP18-SPT21 INTERGENIC REGION.
GN YMR176W OR YMR010.06.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Churcher C.M., Barrell B.G., Raeburn M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: TO MAMMALIAN PROTEIN XE169.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49808; CAA89909.1; -
DR SGD: S0004788; YMR176W.
DR INTERPRO: IPR001606; -
DR INTERPRO: IPR001965; -
DR PFAM: PF01388; ARID.1.
DR PFAM: PF00628; PHD.1.
KW Hypothetical protein.
SQ SEQUENCE 1411 AA: 162700 MW: 18768BD29934D9F9 CRC64;
```

Query Match 6.1%; Score 78.5; DB 1; Length 1411;

Best Local Similarity 19.7%; Pred. No. 35;

Matches 58; Conservative 42; Mismatches 82; Indels 113; Gaps 16;

```
OY 7 EPPKRRKKRDINTIDAVKLLQECKKIIVLGAGVSVSCGIPDFRSRDGIYARLAVDFPD 66
Db 1088 DPKSKISLSD--LEDVLDIEEYRLPI-----QSFSELDYVIRE 1126
OY 67 LPDPOAMFDI-----EYERKDPREFKFAKEIYPGOFPSLCHKFIALSDKE 113
Db 1127 IENAKKMDVNIILYTDNIDKIDELIRKNDKFKFA-----NDFNGSRLDKRLPLASDNS 1181
OY 114 G-----KLLRNTYQNTIDLEQ---VAGIQRILQC----- 139
Db 1182 GSVAKAKQELKFKLMNQHLDQIMQNKKFIELPSIFRCLDKSDKIIPLESCKRQTKYC 1241
OY 140 -----HGSFATASCLICK--YKVDCEAVRGDIENOVPRCPKPADEP-----LAIMKPE 187
Db 1242 FCRRVECTAWVECEICKEMHWDCIS--NGEL-----VPPDDPNVLFVCSICTPP 1290
OY 188 IVFEGENLPEQFHRAKXKDKDEVLDLIYGSLSKVRVALIPSSIPHEVPOIIN 242
Db 1291 CMAV-DNI-----EGVTFLDLDKRIIV--ESLK---LSLIPD-----PPLTKN 1328
```

Search completed: February 16, 2001, 10:52:28
Job time: 531 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:44:01 ; Search time 51.38 seconds

(without alignments)
85.626 Million cell updates/sec

Title: US-09-461-580A-4

Perfect score: 1288

Sequence: 1 VINILSEPPKRRKKRDINTI.....ALIPSPHEVPQILINREP 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	7.5	708	1 US-08-145-681-4	Sequence 4, Appl1
2	97	7.5	708	1 US-08-453-703-4	Sequence 4, Appl1
3	97	7.5	708	2 US-08-456-106-4	Sequence 4, Appl1
4	97	7.5	708	3 US-08-456-108-4	Sequence 4, Appl1
5	82.5	6.4	579	1 US-08-448-196A-8	Sequence 8, Appl1
6	79.5	6.0	329	3 US-09-108-020-48	Sequence 48, Appl1
7	77.5	6.0	2366	1 US-08-480-604A-10	Sequence 10, Appl1
8	77.5	6.0	2366	2 US-08-405-496A-10	Sequence 10, Appl1
9	74	5.7	274	2 US-07-857-224B-30	Sequence 30, Appl1
10	74	5.7	282	1 US-08-318-947A-19	Sequence 19, Appl1
11	74	5.7	282	2 US-08-795-303-19	Sequence 19, Appl1
12	74	5.7	287	2 US-08-874-347-22	Sequence 22, Appl1
13	74	5.7	297	3 US-09-093-522-22	Sequence 22, Appl1
14	72.5	5.6	620	1 US-08-484-105-4	Sequence 4, Appl1
15	72.5	5.6	620	1 US-08-484-106-4	Sequence 4, Appl1
16	72.5	5.6	956	1 US-08-185-232A-2	Sequence 2, Appl1
17	72	5.6	956	1 US-08-416-523-2	Sequence 2, Appl1
18	72	5.6	956	3 US-08-789-478-2	Sequence 2, Appl1
19	71	5.5	1022	1 US-08-271-364A-8	Sequence 8, Appl1
20	71	5.5	1022	2 US-08-222-715B-27	Sequence 27, Appl1
21	70.5	5.5	245	2 US-08-874-138-2	Sequence 2, Appl1
22	70.5	5.5	805	3 US-08-985-526-34	Sequence 34, Appl1
23	70.5	5.5	806	2 US-08-443-861-5	Sequence 5, Appl1
24	70.5	5.5	1367	1 US-07-813-593-4	Sequence 4, Appl1
25	70.5	5.5	1367	1 US-07-977-451-6	Sequence 6, Appl1
26	70.5	5.5	1367	1 US-07-946-507-4	Sequence 4, Appl1
27	70.5	5.5	1367	1 US-08-252-517-6	Sequence 6, Appl1
28	70.5	5.5	1367	1 US-07-906-397A-6	Sequence 6, Appl1

29	70.5	5.5	1367	1 US-08-601-891-6	Sequence 6, Appl1
30	70.5	5.5	1367	2 US-08-443-861-2	Sequence 2, Appl1
31	70.5	5.5	1367	2 US-09-021-324-6	Sequence 6, Appl1
32	70.5	5.5	1367	4 PCT-US92-02750-8	Sequence 8, Appl1
33	70.5	5.5	1367	4 PCT-US92-05401-6	Sequence 6, Appl1
34	70.5	5.5	1367	4 PCT-US92-09893-6	Sequence 8, Appl1
35	70	5.4	259	2 US-08-402-804-8	Sequence 8, Appl1
36	70	5.4	560	3 US-09-132-619-10	Sequence 10, Appl1
37	70	5.4	560	3 US-09-282-803B-10	Sequence 10, Appl1
38	69	5.4	972	3 US-08-335-844A-23	Sequence 23, Appl1
39	69	5.4	2396	1 US-08-157-005-2	Sequence 2, Appl1
40	68.5	5.3	320	2 US-08-846-762-16	Sequence 16, Appl1
41	68.5	5.3	621	3 US-09-059-522-1	Sequence 1, Appl1
42	68.5	5.3	621	3 US-09-382-027-1	Sequence 1, Appl1
43	68.5	5.3	917	2 US-08-588-983-16	Sequence 16, Appl1
44	68.5	5.3	917	2 US-08-588-976-16	Sequence 16, Appl1
45	68	5.3	572	2 US-08-745-934-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-145-681-4
; Sequence 4, Application US/08145681
; Patent No. 5571691
; GENERAL INFORMATION:
; APPLICANT: Connolly, Orla M.
; APPLICANT: Headon, Denis R.
; APPLICANT: O'Malley, Bert W.
; APPLICANT: May, Gregory S.
; TITLE OF INVENTION: Production of Recombinant Lactoferrin
; TITLE OF INVENTION: and Lactoferrin Polypeptides Using cdna Sequences in
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Baker & Botts, L.L.P.
; STREET: 910 Louisiana St
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,681
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGregor, Martin L.
; REGISTRATION NUMBER: 29,329
; REFERENCE/DOCKET NUMBER: 19928-0125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/229/1874
; TELEFAX: 713/229/1522
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; US-08-145-681-4

Query Match 7.5%; Score 97; DB 1; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.021;

Matches	48:	Conservative	26:	Mismatches	57:	Indels	72:	Gaps	9:
QY	81	KDPRPFKRAKITYGQFQPSLCKHFIALSDK-----				EGK-----			115
Db	88	RDPIYLPRVAALITGKRESPOTHYYAAVAVKKGNSNFOLDOLGRRKSCHTGLRSAGMIIP							147
QY	116	---LLRNNTQONIDTLBOVAGIQRIILQCHGSFATASCLIC---				KYKDCAEVAGDIFFNOVV			169
Db	148	MCILRPYLSWTSLSLEPLQG-----				AVAKFFSASCVPCIDRAIYPIMLCOLCGEGENQCA			201
QY	170	PRCPRCPADP-----				LAIMKPEIVFEGENLPRQFHRAMKYDKDEVDL			212
Db	202	-----CSSREPYFGSGAFKCLQDGAADVAFVAKETTVF--				ENLPE-----			KADRDQYEL 248
QY	213	LIVIGS-----				SLKVRVALIPS			230
Db	249	LCILNNSRAPVDAFKKCHLAQVPS				271			

```

1      RESULT      2
2      US-08-453-703-4
3      : Sequence 4, Application US/08453703
4      : Patent No. 5766939
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Conneely, Orla M.
8      : APPLICANT: Heaton, Denis R.
9      : APPLICANT: O'Malley, Bert W.
10     : APPLICANT: May, Gregory S.
11     : TITLE OF INVENTION: Production of Recombinant Lactoferrin
12     : TITLE OF INVENTION: and Lactoferrin Polypeptides Using cDNA Sequences In
13     : NUMBER OF SEQUENCES: 14
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Pennie & Edmonds
16     : STREET: 1155 Avenue of the Americas
17     : CITY: New York
18     : STATE: New York
19     : COUNTRY: USA
20     : ZIP: 10036
21     :
22     : COMPUTER READABLE FORM:
23     : MEDIUM TYPE: Floppy disk
24     : COMPUTER: IBM PC compatible
25     : OPERATING SYSTEM: PC-DOS/MS-DOS
26     : SOFTWARE: Patentin Release #1.0, Version #1.25
27     :
28     : CURRENT APPLICATION DATA:
29     : APPLICATION NUMBER: US/08/453,703
30     : FILING DATE: Concurrently herewith
31     : CLASSIFICATION: 435
32     :
33     : PRIOR APPLICATION DATA:
34     : APPLICATION NUMBER: 08/145,681
35     : FILING DATE: October 28, 1993
36     : ATTORNEY/AGENT INFORMATION:
37     : NAME: Albert P. Halluin
38     : REGISTRATION NUMBER: 25,227
39     : REFERENCE/DOCKET NUMBER: 8206-024
40     : TELECOMMUNICATION INFORMATION:
41     : TELEPHONE: 415-854-3660
42     : TELEFAX: 415-854-3694
43     : INFORMATION FOR SEQ ID NO: 4:
44     : SEQUENCE CHARACTERISTICS:
45     : LENGTH: 708 amino acids
46     : TYPE: amino acid
47     : TOPOLOGY: linear
48     : MOLECULE TYPE: protein
49     : HYPOTHEetical: YES
50     : ANTI-SENSE: NO
51     : ORIGINAL SOURCE:
52     : ORGANISM: Bos taurus
53     :
54     : US-08-453-703-4

```

Query Match	7.58; Score 97; DB 1; Length 708;
Best Local Similarity	23.68; Pred. No. 0.021;

Matches	48;	Conservative	26;	Mismatches	57;	Indels	72;	Gaps	9;
QY	81	KDPRPEFFKATITYGQFQPSLCHKFI	ALSDK-----	EGK-----					115
Db	88	RDPIYLRLPVAAALIEYGTKESSPQTH	YAYAVAVKKGNSFNOLDLOGRS	CHTGLGRSAGWIIP					147
QY	116	---LLNNYTNQITLROVAGIQRIILQCH	GSFATSLC---	KYVDEAVNGDIF	FNQVY				169
Db	148	MCILRPYLSTWSTESLEPLQG-----	AVAKFFSASVPCICIDRQATFNL	COLCKGEGENQCA					201
QY	170	PRCPGCPADP-----	LAIKKPEIYFEGENLPEQ	OFHRAMKDYKDEVDL					212
Db	202	-----CSSREPYFGSGAFKCLQD	GAGDAVAFVAKETTVF--	ENLPE-----	KADRQDEL				248
QY	213	LIVIGS-----	SLKVRPAVALIPS						230
Db	249	LCINNSRAIVDAFKEC	CHLAQVPS						271

```

1      RESULT      3
2      US-08-456-106-4
3      ; Sequence 4, Application US/08456106
4      ; Patent No. 5849881
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Conneely, Orla M.
7      ; APPLICANT: Heaton, Denis R.
8      ; APPLICANT: O'Malley, Bert W.
9      ; APPLICANT: May, Gregory S.
10     ; TITLE OF INVENTION: Production of Recombinant Lactoferrin
11     ; TITLE OF INVENTION: and Lactoferrin Polypeptides using cDNA Sequences In
12     ; NUMBER OF SEQUENCES: 14
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSEE: Pennie & Edmonds
15     ; STREET: 1155 Avenue of the Americas
16     ; CITY: New York
17     ; STATE: New York
18     ; COUNTRY: USA
19     ; ZIP: 10036
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Floppy disk
22     ; COMPUTER: IBM PC compatible
23     ; OPERATING SYSTEM: PC-DOS/MS-DOS
24     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/08/456.106
27     ; FILING DATE: Concurrently herewith
28     ; CLASSIFICATION: 435
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: 08/145.681
31     ; FILING DATE: October 28, 1993
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: Albert P. Halluin
34     ; REGISTRATION NUMBER: 25,227
35     ; REFERENCE/DOCKET NUMBER: 8206-025
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: 415-854-3660
38     ; TELEFAX: 415-854-3694
39     ; INFORMATION FOR SEQ ID NO: 4:
40     ; SEQUENCE CHARACTERISTICS:
41     ; LENGTH: 708 amino acids
42     ; TYPE: amino acid
43     ; TOPOLOGY: linear
44     ; MOLECULE TYPE: protein
45     ; HYPOTHEICAL: YES
46     ; ANTI-SENSE: NO
47     ; ORIGINAL SOURCE:
48     ; ORGANISM: Bos taurus
49     ; US-08-456-106-4

```

Query Match	7.58; Score 97; DB 2; Length 708;
Best Local Similarity	23.68; Pred. No. 0.021;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDRPFKFAKEIYRGFQSLCKHFIALSDK-----EGR----- 115
DB 88 RDPKLRPVAAEITGESQTHYAAVAVKGSNFOLDOLQGRKSGHTGLGRSAGNIIP 147
QY 116 -LLRNTQNIIDLEQVAGIQRILQCHGSPATASCLIC-----KKVDCNAVGRDIFNQV 169
DB 148 MGIIRPLSTWESLEPIQG-----AAKFFSASCVCIDRQAVPNLCQLCKGGENQCA 201
QY 170 PRCPRCADPE-----LAIMKEIVFEGENLPEQFRAMKYDKDEVLD 212
DB 202 -----CSSRPEYEGYSGAFKCLQDGAGDAVAFKETTVE--ENLPE-----KADRDQYEL 248
QY 213 LIVIGS-----SLKVRVALIPS 230
DB 249 LCLNNSRAPVDAFKECHLAQVPS 271

RESULT 4
US-08-456-108-4
; Sequence 4, Application US/08456108
; Patent No. 6100054
; GENERAL INFORMATION:
; APPLICANT: Conneely, Orla M.
; APPLICANT: Headon, Denis R.
; APPLICANT: O'Malley, Bert W.
; TITLE OF INVENTION: Production of Recombinant Lactoferrin
; TITLE OF INVENTION: and Lactoferrin Polypeptides Using cDNA Sequences in
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,108
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/145,681
; FILING DATE: October 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8206-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; US-08-456-108-4

Query Match 7.5%; Score 97; DB 3; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.021;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDRPFKFAKEIYRGFQSLCKHFIALSDK-----EGR----- 115
DB 88 RDPKLRPVAAEITGESQTHYAAVAVKGSNFOLDOLQGRKSGHTGLGRSAGNIIP 147
QY 116 -LLRNTQNIIDLEQVAGIQRILQCHGSPATASCLIC-----KKVDCNAVGRDIFNQV 169
DB 148 MGIIRPLSTWESLEPIQG-----AAKFFSASCVCIDRQAVPNLCQLCKGGENQCA 201
QY 170 PRCPRCADPE-----LAIMKEIVFEGENLPEQFRAMKYDKDEVLD 212
DB 202 -----CSSRPEYEGYSGAFKCLQDGAGDAVAFKETTVE--ENLPE-----KADRDQYEL 248
QY 213 LIVIGS-----SLKVRVALIPS 230
DB 249 LCLNNSRAPVDAFKECHLAQVPS 271

RESULT 5
US-08-448-196A-8
; Sequence 8, Application US/08448196A
; Patent No. 5780594
; GENERAL INFORMATION:
; APPLICANT: CARTER, DANIEL C.
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN FRAGMENTS
; TITLE OF INVENTION: CONTAINING SPECIFIC BINDING REGIONS OF SERUM ALBUMIN OR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NASA
; STREET: MARSHALL SPACE FLIGHT CENTER
; CITY: HUNTSVILLE
; STATE: ALABAMA
; COUNTRY: USA
; ZIP: 35812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,196A
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: BROAD JR., ROBERT L.
; REGISTRATION NUMBER: 18,757
; REFERENCE/DOCKET NUMBER: XX/MFS-28402-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 205-544-0021
; TELEFAX: 205-544-0258
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 579 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-448-196A-8

Query Match 6.4%; Score 82.5; DB 1; Length 579;
Best Local Similarity 22.8%; Pred. No. 0.69;
Matches 50; Conservative 35; Mismatches 81; Indels 53; Gaps 11;

QY 45 CGIPDFSRDGIYARLAVDPDLPQAMFDIEYFRDPRFFKFAK-----EITYGQGFQ 99
DB 276 CNLP---LLERTYCIIVLENDVP-AELSKPITFTEDPHVCQYVANKKSFLEISPMQSQ 131
QY 100 --PSLCHKFIALSDKEKGLRN-----YTONIDRL-----EQVAGIQR---ILQ 138

Db 332 ETPELSEDFLOSAKEYESLKKCFSDNPECYDGDAREFNENKERRAYLKQNCIDLH 391
Oy 139 CHGSFATASCLICKYKVCEAVRGDIFNQVYPRCPDPADEPLAMKREIYFEGN---L 195
Db 392 EHGELFENELLIRI-----TKMPQV-SDELLIGIANOMADIGHCACV 435
Oy 196 PEQFHRAMKYDKEVDLLIVIGSSLSKVRPVALIPSSIPH 234
Db 436 PEN---QRMPCAEGLDFTLLICKMCEBQKTFEINNVAH 470

RESULT 6
US-09-108-020-48
Sequence 48, Application US/09108020A
Patent No. 6143561
GENERAL INFORMATION:
APPLICANT: Randall, Douglas D.
APPLICANT: Johnston, Mark L.
APPLICANT: Miernyk, Jan A.
APPLICANT: Luethy, Michael H.
APPLICANT: Mooney, Brian P.
TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
FILE REFERENCE: UMO 1482
CURRENT FILING DATE: 1998-06-30
EARLIER FILING DATE: 1997-06-30
EARLIER FILING DATE: 1997-06-30
EARLIER FILING DATE: 1997-06-30
EARLIER FILING DATE: 1997-08-01
EARLIER FILING DATE: 1997-08-01
EARLIER FILING DATE: 1998-03-02
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 329
TYPE: PRT
ORGANISM: M. capricolum
US-09-108-020-48

Query Match 6.2%; Score 79.5; DB 3; Length 329;
Best Local Similarity 23.1%; Pred. No. 0.66;
Matches 51; Conservative 31; Mismatches 100; Indels 39; Gaps 10;
Oy 16 DINTEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARLADFP-----DLFDP 70
Db 6 NIKAVTDALDCAMQRPVIVFEGEDVTEGGV--FRATQG--LAVKFGNDRCFNADIS 59
Oy 71 QAMF---DIEYFRKDPREFKFAKIYPGQPSLCHKFIKILSDSEGKLLNRYTNIDTL 127
Db 60 EAMFPGVGLGMAAMGKRVLEMO---FESGLASLQNIFTNISRNRNTRGKTYTAPVIR 116
Oy 128 EOVAGIQRILOCHGSFATASCLICKYKVCEAVRGDIFN-QVPRCPDPADEP-----LA 182
Db 117 MPMGSGIALEHNSE-----ALEAVYAHIRGQYV--CPSTPYDTKLLILAA 161
Oy 183 IMKPEIYFEGNLPQGFHRAMKYDKEVDLLIVIGSSLSKVR 223
Db 162 IDSPDPVIVE--PTKLYRAFKQEVDPDEHYIVPIGEGYKIQ 200

RESULT 7
US-08-480-604A-10
Sequence 10, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 6.0%; Score 77.5; DB 1; Length 2366;
Best Local Similarity 17.8%; Pred. No. 20;
Matches 54; Conservative 50; Mismatches 125; Indels 75; Gaps 8;
Oy 2 INIIS-----EPPKRRKRKDINTIEDAVKLLQECKKIIVLTGAGVSVSCGIDTF 50
Db 826 INVINSIDTQIVERIEIEKAKMLTSDSINVIKDEFLEI-----SIDALCDL 872
Oy 51 RSPDGIYARLAVDPDLPDPAQMFIEYFRKDP--FEKFAKIYPGQPSLCHKFIA 108
Db 873 KOQNELEDSHFISFEDISFETDEGFSIRINKETGESIVETEKITF--SEYNNHTEEISK 931
Oy 109 LSDK-----EGKLLR---NTYONIDTLEQVAGIQRILQCHGSFATASCLICKYKVCEA 159
Db 932 IKGTIFDVNGKLVKQVLDTHHEVNTLNAFPIQSLIEYSSKESLSLNVAKKQVYA 991
Oy 160 -----VRDGINQVYPRCPDPADEPLAMKREIYFEGNLPQGFHR 201
Db 992 QLFSTGLNTTDAKVVELVSTALDETIDLPITLSEGLPIATIIDGVSLGAAIKELSET 1051

```
Query Match      6.0% ; Score 77.5 ; DB 2 ; Length 2366 ;
Best Local Similarity 17.8% ; Pred. No. 20 ;
Matches 54 ; Conservative 50 ; Mismatches 125 ; Indels 75 ; Gaps 8

Qy 2 INILS-----EPPRRKRKRDINITEAVKLQCEKKIIVLTGAGVSVCIDPF 50
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 826 INVISINIDTQVIEERLEAKKNLTSISINIKDEPKILE-----SISDMLCOL 872
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy 51 RSRDDIYARLVAFDFDLPPQAMFEDIERKDRP--EFKFAKEYIGGFQPSLCHKFTA 108
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
```

```

1 RESULT 9
2 US-07-857-224B-30
3 ; Sequence 30 Application US/07857224B
4 ; Patent No. 5958784
5 ;
6 ; GENERAL INFORMATION:
7 ; APPLICANT: Benner, Steven A.
8 ; TITLE OF INVENTION: Predicting Folded Structures of Proteins
9 ; NUMBER OF SEQUENCES: 114
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Steven A. Benner
12 ; STREET: Hadlaubstrasse 151
13 ; CITY: Zurich
14 ; STATE: none
15 ; COUNTRY: Switzerland
16 ;
17 ; ZIP: (note: this is an international post code) CH-8092
18 ;
19 ; COMPUTER READABLE FORM:
20 ; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
21 ; COMPUTER: Apple Macintosh
22 ; OPERATING SYSTEM: Macintosh 7.0
23 ; SOFTWARE: Microsoft Word
24 ;
25 ; CURRENT APPLICATION DATA:
26 ; APPLICATION NUMBER: US/07/857,224B
27 ; FILING DATE: 03/25/92
28 ;
29 ; CLASSIFICATION: 436
30 ;
31 ; PRIOR APPLICATION DATA: none
32 ; TELECOMMUNICATION INFORMATION:
33 ; TELEPHONE: (International) 41 1 632 2830
34 ; TELEFAX: (International) 41 1 262 2437
35 ;
36 ; TELEX: none
37 ;
38 ; INFORMATION FOR SEQ ID NO: 30:
39 ;
40 ; SEQUENCE CHARACTERISTICS:
41 ;
42 ; LENGTH: 274
43 ; TYPE: amino acid
44 ; TOPOLOGY: linear
45 ;
46 ; MOLECULE TYPE: protein
47 ;
48 ; DESCRIPTION: protein
49 ;
50 ; ORIGINAL SOURCE:
51 ; ORGANISM: Schizosaccharomyces pombe
52 ; FEATURE: Protein kinase; Table 8 Column 34
53 ; PUBLICATION INFORMATION:
54 ;
55 ; AUTHORS:
56 ;
57 ; AUTHORS: Hanks, S. K.
58 ; AUTHORS: Quinn, A. M.
59 ; AUTHORS: Hunter, T.
60 ; TITLE: The protein kinase family
61 ; JOURNAL: Science
62 ; VOLUME: 241
63 ;
64 ; PAGES: 42-52
65 ;
66 ; DATE: 1986
67 ;
68 ; US-07-857-224B-30

```

Query Match 5.7%; Score 74; DB 2; Length 274;
Best Local Similarity 29.0%; Pred. NO. 2.1;

Matches 36; Conservative 12; Mismatches 38; Indels 38; Gaps 7;

QY 82 DRRPEFKAKIYPGQFSPICH--KFI-----ALSDKEGKL-----L 117
DB 104 DPLVQKFTYQLVNG---VNFCHSRRIIHRDLKPONLLIDEGNLKADFGIARSGVPL 160

QY 118 RNYTQNTIDTLEOVAGIQRILQCHGSFATASCLICKYKVDCAVNGDIFNQVPRCPA 177
DB 161 RNYTHEIYTLWYRA--PEVLLGSRHYSTG-----VDIMSV-GCIFAEMIRSRPLPFG 209

QY 178 DEPL 181
DB 210 DSEI 213

RESULT 10
US-08-318-947A-19
Sequence 19, Application US/08318947A
Patent No. 5798245
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
TITLE OF INVENTION: TIA-1 BINDING PROTEINS AND ISOLATED
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrie, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, NW Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,947A
FILING DATE: 06-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/133,530
FILING DATE: 07-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mack, Susan J.
REGISTRATION NUMBER: 30,951
REFERENCE/DOCKET NUMBER: A6462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)293-7060
TELEFAX: (202)293-2920
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-318-947A-19

Query Match 5.7%; Score 74; DB 1; Length 282;
Best Local Similarity 29.0%; Pred. No. 2.2;
Matches 36; Conservative 12; Mismatches 38; Indels 38; Gaps 7;

QY 82 DRRPEFKAKIYPGQFSPICH--KFI-----ALSDKEGKL-----L 117
DB 107 DPLVQKFTYQLVNG---VNFCHSRRIIHRDLKPONLLIDEGNLKADFGIARSGVPL 163

QY 118 RNYTQNTIDTLEOVAGIQRILQCHGSFATASCLICKYKVDCAVNGDIFNQVPRCPA 177
DB 164 RNYTHEIYTLWYRA--PEVLLGSRHYSTG-----VDIMSV-GCIFAEMIRSRPLPFG 212

QY 178 DEPL 181
DB 213 DSEI 216

RESULT 11
US-08-795-303-19
Sequence 19, Application US/08795303
Patent No. 5948656
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
TITLE OF INVENTION: TIA-1 BINDING PROTEINS AND ISOLATED
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrie, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, NW Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,303
FILING DATE: 04-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/318,947
FILING DATE: 06-OCT-1994
APPLICATION NUMBER: 08/133,530
FILING DATE: 07-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mack, Susan J.
REGISTRATION NUMBER: 30,951
REFERENCE/DOCKET NUMBER: A6462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)293-7060
TELEFAX: (202)293-2920
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-303-19

Query Match 5.7%; Score 74; DB 2; Length 282;
Best Local Similarity 29.0%; Pred. No. 2.2;
Matches 36; Conservative 12; Mismatches 38; Indels 38; Gaps 7;

QY 82 DRRPEFKAKIYPGQFSPICH--KFI-----ALSDKEGKL-----L 117
DB 107 DPLVQKFTYQLVNG---VNFCHSRRIIHRDLKPONLLIDEGNLKADFGIARSGVPL 163

QY 118 RNYTQNTIDTLEOVAGIQRILQCHGSFATASCLICKYKVDCAVNGDIFNQVPRCPA 177
DB 164 RNYTHEIYTLWYRA--PEVLLGSRHYSTG-----VDIMSV-GCIFAEMIRSRPLPFG 212

QY 178 DEPL 181
DB 213 DSEI 216

RESULT 12
US-08-874-347-22
Sequence 22, Application US/08874347

```
Patent No. 5863741
GENERAL INFORMATION:
APPLICANT: Limper, Andrew H.
APPLICANT: Leof, Edward B.
APPLICANT: Thomas, Charles F.
APPLICANT: Gustafson, Michael P.
TITLE OF INVENTION: CDC2 PROTEIN KINASE FROM PNEUMOCYSTIS
TITLE OF INVENTION: CARINI
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C., P.A.
STREET: 60 South Sixth Street, Suite 3300
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874,347
FILING DATE: 13-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ellinger, Mark S.
REGISTRATION NUMBER: 34,812
REFERENCE/DOCKET NUMBER: 07039/055001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-335-5070
TELEFAX: 612-288-9696
TELEX:
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-874-347-22

Query Match          5.7%; Score 74; DB 2; Length 297;
Best Local Similarity 29.0%; Pred. No. 2.4;
Matches 36; Conservative 12; Mismatches 38; Indels 38; Gaps 7;

QY 82 DRRPFFFAKEIYPGQFQPSLCH--KFI-----ALSDKEGKL-----L 117
DB 107 DPLRYQKFTYQLVNG---VNFCHSRRIIHRDKPQNLIDKEGKLADGLARSGVPL 163
QY 118 RNTQNTIDLEQVAGIRIIOCHGSFATASCLICKRYVDCEAVGDIENOVVPRCPGPA 177
DB 164 RNTYHEIIVTLMYRA--DEVLLGSRHYSTG-----VDIMSV-GCIFAEMIRSRPLPG 212
QY 178 DEPL 181
DB 213 DSET 216

RESULT 13
US-09-093-522-22
Sequence 22, Application US/09093522
GENERAL INFORMATION:
APPLICANT: Limper, Andrew H.
APPLICANT: Leof, Edward B.
APPLICANT: Thomas, Charles F.
APPLICANT: Gustafson, Michael P.
TITLE OF INVENTION: CDC2 PROTEIN KINASE FROM PNEUMOCYSTIS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Fish & Richardson P.C., P.A.
STREET: 60 South Sixth Street, Suite 3300
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/093,522
FILING DATE: 08-JUN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/874,347
FILING DATE: 13-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ellinger, Mark S.
REGISTRATION NUMBER: 34,812
REFERENCE/DOCKET NUMBER: 07039/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-335-5070
TELEFAX: 612-288-9696
TELEX:
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-093-522-22
```

```
Query Match          5.7%; Score 74; DB 3; Length 297;
Best Local Similarity 29.0%; Pred. No. 2.4;
Matches 36; Conservative 12; Mismatches 38; Indels 38; Gaps 7;

QY 82 DRRPFFFAKEIYPGQFQPSLCH--KFI-----ALSDKEGKL-----L 117
DB 107 DPLRYQKFTYQLVNG---VNFCHSRRIIHRDKPQNLIDKEGKLADGLARSGVPL 163
QY 118 RNTQNTIDLEQVAGIRIIOCHGSFATASCLICKRYVDCEAVGDIENOVVPRCPGPA 177
DB 164 RNTYHEIIVTLMYRA--DEVLLGSRHYSTG-----VDIMSV-GCIFAEMIRSRPLPG 212
QY 178 DEPL 181
DB 213 DSET 216

RESULT 14
US-08-484-105-4
Sequence 4, Application US/08484105
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P.
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: KINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J.
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J.
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:49:13 ; Search time 220.15 Seconds
(without alignments)
169.331 Million cell updates/sec

Title: US-09-461-580A-4
Perfect score: 1288
Sequence: 1 VINILSEPPRRKKRDKINTI.....ALIPSSIPHEVQILINREP 245

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 936741 seqs, 152156132 residues
Total number of hits satisfying chosen parameters: 936741

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents-AA:*

```
1: /cgn2_6/ptodata/2/paa/PCFUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	100.0	245	US-09-461-580A-4	Sequence 4, Appl1
2	1288	100.0	327	US-09-461-580A-12	Sequence 12, Appl1
3	1288	100.0	737	US-09-461-580A-1	Sequence 1, Appl1
4	1288	100.0	737	US-09-461-580A-26	Sequence 26, Appl1
5	883.5	68.6	823	US-09-619-049-576	Sequence 576, Appl1

6	883.5	68.6	823	US-60-167-217-10885	Sequence 10885, A
7	883.5	68.6	823	US-60-171-627-915	Sequence 915, App
8	883.5	68.6	823	US-60-173-464-8827	Sequence 8827, Ap
9	883.5	68.6	823	US-60-191-637-10901	Sequence 10901, A
10	883.5	68.6	823	US-60-191-681-8549	Sequence 8549, Ap
11	747	58.0	286	US-09-270-767-44940	Sequence 44940, A
12	682	53.0	128	US-09-461-580A-9	Sequence 9, Appl1
13	589.5	45.8	339	US-09-248-796-19599	Sequence 19599, A
14	589.5	45.8	339	US-60-096-409-19599	Sequence 19599, A
15	589.5	45.8	339	US-09-461-580A-19	Sequence 19, Appl1
16	530	41.1	389	US-60-169-629-517	Sequence 3, Appl1
17	530	41.1	389	US-60-167-217-10557	Sequence 517, App
18	530	41.1	389	US-09-461-580A-2	Sequence 2, Appl1
19	517	40.1	272	US-09-461-580A-11	Sequence 11, Appl1
20	517	40.1	336	US-09-248-796-20918	Sequence 20918, A
21	512	39.8	313	US-60-096-409-20918	Sequence 20918, A
22	512	39.8	313	US-60-167-217-10557	Sequence 10557, A
23	489	38.0	355	US-60-173-464-8533	Sequence 8533, Ap
24	489	38.0	355	US-60-117-905-57	Sequence 57, Appl1
25	470	36.5	381	US-60-096-827-17	Sequence 17, Appl1
26	468.5	36.4	257	US-09-248-796-19606	Sequence 19606, A
27	449	34.9	284	US-60-096-409-19606	Sequence 19606, A
28	449	34.9	284	PCT-US00-02237-9	Sequence 9, Appl1
29	441.5	34.3	230	US-60-117-905-9	Sequence 9, Appl1
30	441.5	34.3	230	US-09-417-507-38938	Sequence 38938, A
31	422	32.8	358	US-09-461-580A-35	Sequence 35, Appl1
32	406	31.5	183	US-09-461-580A-20	Sequence 20, Appl1
33	367.5	28.5	232	US-09-461-580A-10	Sequence 10, Appl1
34	312.5	24.3	107	US-09-461-580A-10	Sequence 10, Appl1
35	304	23.6	128	US-09-461-580A-10	Sequence 10, Appl1
36	287.5	22.3	445	US-60-117-905-70	Sequence 70, Appl1
37	285	22.1	248	US-09-248-796-19601	Sequence 19601, A
38	285	22.1	488	US-60-096-409-19601	Sequence 19601, A
39	268.5	20.8	248	US-09-450-965-6406	Sequence 6406, Ap
40	267.5	20.8	223	PCT-US99-29950-188	Sequence 188, App
41	267.5	20.8	223	US-09-591-316-189	Sequence 189, App
42	266.5	20.7	107	US-09-461-580A-16	Sequence 16, Appl1
43	264	20.5	270	US-09-252-991A-19648	Sequence 19648, A
44	258.5	20.1	253	US-09-611-529-6441	Sequence 6441, Ap
45	246	19.1	106	US-09-461-580A-15	Sequence 15, Appl1

ALIGNMENTS

```
RESULT 1
US-09-461-580A-4
; Sequence 4, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-4
```

Query Match 100.0%; Score 1288; DB 18; Length 245;
Best Local Similarity 100.0%; Pred. No. 5.8e-136;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VINILSEPPRRKKRDKINTIEDAVKLLQCKKTIIVTGACVSCGIPDRSRDGIARL 60
|||||

D6	1	VINILSEPPKKKKKRDINTTIEDAKLLQOECKIIVLTGAGVSVSCGIPDRSRDGIARL	60
QY	61	AVDEPDLPDQAMNDIEYFRKDRPPFFKFAKEIYPGQFQFSLCHKFTALSDKSGKLLRNY	120
D6	61	AVDEPDLPDQAMPDIEYFRKDRPPFFKFAKEIYPGQFQFSLCHKFTALSDKSGKLLRNY	120
QY	121	TQNTDTEQVAGIQRILQCHGSFATASCLCKYKVDCEAVRGDIFNNOVYRCRCPADEP	180
D6	121	TQNTDTEQVAGIQRILQCHGSFATASCLCKYKVDCEAVRGDIFNNOVYRCRCPADEP	180
QY	181	LAIKKPELVEFGENLPEQFHRAMKYDKDEVDLLIVIGSSLKVRVALIPSSIEHEVPQIL	240
D6	181	LAIKKPELVEFGENLPEQFHRAMKYDKDEVDLLIVIGSSLKVRVALIPSSIEHEVPQIL	240
QY	241	INREP 245	
D6	241	INREP 245	

```

RESULT      2
US-09-461-580A-12
: Sequence 12, Application US/09461580A
: GENERAL INFORMATION:
: APPLICANT: Guarante, Leonard
: APPLICANT: Imai, Shin-ichiro
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASED
: TITLE OF INVENTION: LIPESPAN
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12
: LENGTH: 327
: TYPE: PRT
: ORGANISM: Mus musculus
: US-09-461-580A-12

```

	Query Match	100.0%	Score 1288:	DB 18;	Length 327;
	Best Local Similarity	100.0%;	Pred.	No. 9e-136;	
Matches	245;	Conservative	0;	Mismatches	0;
OY	1 VINILSEPPKRRKRNDITIEDAVKLLOECKKIIVLTGAGVSVSCGIPIDFRSRDIYARL	60			
Sb	1 VINIASEPKKKRKNDITIEDAVVLLOECKKIIVLTGAGVSVCIGIDFRSRDGIYARL	60			
OY	. 61 AVDPDLDPDPMADIEYFKRDPRFEFFFAKEIIYPGGFQPSLCHEFIALSDKEGKLIRNY	120			
Sb	61 AVDPDDLDPDPMADIEYFKRDPRFEFFFAKEIYPGGCFQPSLSCHKFIALSDEGEKLIRNY	120			
OY	121 TONIDLTEOVAGIORRIIOCHGSFATASCLICKKYVDCAVNGDIJFNQVPKPCRPCEPADP	180			
Sb	121 TONIDLTEOVAGIORRIIOCHGSFATASCLICKKYVDCAVNGDIJFNQVPKPCRPCEPADP	180			
OY	181 LAIKPEIIVEFGENLPEDGFHRAMKYDKDEVDLLITYGSSLKVRPALIPSPISPHVEPOLI	240			
Sb	181 LAIKPEIIVEFGENLPEDGFHRAMKYDKDEVDLLITYGSSLKVRPALIPSPISPHVEPOLI	240			
OY	241 INREP 245				
Sb	241 INREP 245				

RESULT 3
US-09-461-580A-1
; Sequence 1, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT : Guarente, Leonard
; APPLICANT : Imai, Shin-Ichiro
APPLICANT : Armstrong, Christopher

```

: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
: TITLE OF INVENTION: LIFESPAN
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ. ID NOS: 35
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 737
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-461-580A-1

```

Query Match	100.0%;	Score 1288;	DB 18;	Length 737;
Best Local Similarity	100.0%;	Pred. No. 3.1e-135;		
Matches 245; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	VINILSEPPKRRKRDIINTIEDAKLLQCEKKIIVLGGAGSVSCGJIPDRSRGIIARL	60
Db	216	VINILSEPPKRRKRDIINTIEDAKLLQCEKKIIVLGGAGSVSCGJIPDRSRGIIARL	275
QY	61	AVDEPDLDPQAMFDIEYFKDPPFFKFAKEIYPGQFQPSLCHKFTALSDKEGKLLRNY	120
Db	276	AVDEPDLDPQAMFDIEYFKDPPFFKFAKEIYPGQFQPSLCHKFTALSDKEGKLLRNY	335
QY	121	TQNIIDLTEQVAVGIORILIOCHGSFATASCLICKYKVDEAVNRGDIFFNOVYPRCPADPEP	180
Db	336	TQNIIDLTEQVAVGIORILIOCHGSFATASCLICKYKVDEAVNRGDIFFNOVYPRCPADPEP	335
QY	181	LAIKKPELVEFGEMLPQOFHRAMKYDKDEVDLLIVIGSSLKVRVALIIPSSIPHEVPQIL	240
Db	396	LAIKKPELVEFGEMLPQOFHRAMKYDKDEVDLLIVIGSSLKVRVALIIPSSIPHEVPQIL	455
QY	241	INREP 245	
Db	456	INREP 460	

```

RESULT      4
US-09-461-580A-26
: Sequence 26, Application US/09461580A
:
: GENERAL INFORMATION:
:
: APPLICANT: Guarente, Leonard
:
: APPLICANT: Imai, Shin-ichiro
:
: APPLICANT: Armstrong, Christopher
:
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
:
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
:
: TITLE OF INVENTION: LIFESPAN
:
: FILE REFERENCE: 0050.1618-000
:
: CURRENT APPLICATION NUMBER: US/09/461,580A
:
: CURRENT FILING DATE: 1999-12-15
:
: NUMBER OF SEQ ID NOS: 35
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 26
:
: LENGTH: 737
:
: TYPE: PR1
:
: ORGANISM: Mus musculus
:
: US-09-461-580A-26

```

Query Match	100.0%;	Score 1288;	DB 18;	Length 737;
Best Local Similarity	100.0%;	Pred. No. 3.1e-135;		
Matches 245;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	VIIISPPRRKKRRKDINTEDAVKLLQEQKKIIVLTGAQSVSCGIPDRSRDGIYARL	60
Db	216	VIIISLPPRRKKRRKDINTEDAVKLLQEQKKIIVLTGAQSVSCGIPDRSRDGIYARL	275
Qy	61	AVDFPDLPPQAMFDIEYFRKDPPEFFKAEKYEI PGQFQPSLCHKETALSDKEGKLLRNY	120
276	AVDFPDLPPQAMFDIEYFRKDPPEFFKAEKYEI PGQFQPSLCHKETALSDKEGKLLRNY	335	

Qy	121	TONIDTLEOVAGIORITLOCHGSEFAATSCICIKKXKDCAEVRGDIENOVVPRCPRADEP	180
Db	336	TONIDTLEOVAGIORITLOCHGSEFAATSCICIKKXKDCAEVRGDIENOVVPRCPRADEP	395
Qy	161	LAIMKPEIVFEGENLPEOPHRAMKXDKDVEDLLIYGSSLKVRPALIPSSIPIHEVPOLL	240
Db	396	LAIMKPEIVFEGENLPEOPHRAMKXDKDVEDLLIYGSSLKVRPALIPSSIPIHEVPOLL	455
Qy	241	INREP 245	
Db	456	INREP 460	

```

RESULT      5
US-09-619-049-576
: Sequence 576: Application US/09619049
: GENERAL INFORMATION:
: APPLICANT: VANDELL, MARK
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
: TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
: TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSCTICIDAL
: TITLE OF INVENTION: TARGETS
: FILE REFERENCE: C1000735
: CURRENT APPLICATION NUMBER: US/09/619, 049
: CURRENT FILING DATE: 2000-07-18
: PRIOR APPLICATION NUMBER: 60/111,590
: PRIOR FILING DATE: 1999-12-23
: PRIOR APPLICATION NUMBER: 60/111,627
: PRIOR FILING DATE: 1999-12-23
: PRIOR APPLICATION NUMBER: 60/175,763
: PRIOR FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: 60/175,685
: PRIOR FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: 60/166,663
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: 60/187,241
: PRIOR FILING DATE: 2000-03-03
: NUMBER OF SEQ ID NOS: 1533
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 576
: LENGTH: 823
: TYPE: PRT
: ORGANISM: DROSOPHILA
US-09-619-049-576

```

[illegible]

```

US-60-167-217-10885
; Sequence 10885, Application US/60167217
;
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1000152
;
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
;
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10885
;
; LENGTH: 823
;
; TYPE: prt
;
; ORGANISM: Drosophila
;
US-60-167-217-10885

```

[illegible]

```

07      RESULT          7
US-60-171-627-915
; Sequence 915, Application US/60171627
; GENERAL INFORMATION:
; APPLICANT: Yandell, Mark
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00177
; CURRENT APPLICATION NUMBER: US/60/171,627
; CURRENT FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-171-627-915

Query Match           68.6%; Score 883.5; DB 22; Length 823;
Best Local Similarity 63.6%; Pred. No. 9.9e+90;
Matches 166; Conservative 38; Mismatches 38; Indels 19; Gaps 2

0Y      3 NILSEPPKRRKKRDINTIEDAVKLQEOCKKIIVLNGAGVSCGIPDFDSRGIYARLAV 62
       :|||::||: ||:||: |::: ||||||||| ||||| |||||
Db     194 HILNEPKRRNKLASVNTPEVDVISLVKSKSIITVLIGAGVSCGIPDFSTNGIARLAA 253

0Y      63 DEPDLPDPOAMDIEYFRKKDPREFRFAKEIYGQFQPSLCHKFTALSDRECKLLRNYYQ 122
```

Db	254	DEPDLPRDQAFNDLNYFRNDPRPRYKFAREIYBPSEFPSPCHRIKMLEKTKILIRNYTQ	3133
QY	123	NIDTLEOVAGIORILQCHGSFATASCILCKKYDCAEVRGDIENOVVPRCPRCP-----	176
Db	314	NIDTLERAVAGIORVIECHGSSTSCCKCRKCANADLRADIFRQRIIPVCPQCPQNKES	373
QY	177	-----ADEPL-----AIKPEIYFEGENLPRQFRANKYDKDEVDLLIVIGSSIKVR	223
Db	374	VDASAVAYEEELRQIVENGIMKPRIVFEGESLPRBYHTVAITDKDVCDDLIVIGSSIKVR	433
QY	224	PVALIPSSIPHEVPQILINRE	244
Db	434	PVAHPPSSIPATVPQILINRE	454

```

RESULT      8 44-8827
US-60-173-464-8827
: Sequence 8827, Application US/60173464
:
: GENERAL INFORMATION:
:
: APPLICANT: LI, Peter W.D.
:
: TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR, PROTEINS AND USES
:
: TITLE OF INVENTION: THEREOF
:
: FILE REFERENCE: CLO00173
:
: CURRENT APPLICATION NUMBER: US/60/173,464
:
: CURRENT FILING DATE: 1999-12-29
:
: NUMBER OF SEQ ID NOS: 30269
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 8827
:
: LENGTH: 823
:
: TYPE: PRT
:
: ORGANISM: Drosophila
:
: US-60-173-464-8827

```

	Query Match	68.6%	Score 883.5	DB 22	Length 823
	Best Local Similarity	63.6%	Pred. No. 9.9e-90		
	Matches 166	Conservative 38	Mismatches 36	Indels 19	Gaps 2
Qy	3.	NILSEPPRRKKRKDDINTEDAVKLLIOECKKIITVLGAGVSCGIPDFFSRDGIYARLAV	62		
Db	194	HLNLEPKRRNNKASVANTFEDDVISLWKKSKKIIIVLGAGVSCGIPDFFSTNGIYARLH	253		
Qy	63	DEPDLPDQAMDEIYFRKDPPEFKFAKEIYGGOFPSLGHFTALSRECKLLRNNTQ	122		
Db	254	DEPDLPDQAMDEIYFRKDPPEFKFAKEIYGGOFPSRPHRIKMLETKGKLLRNNTQ	313		
Qy	123	NIDTLEQVAGIORILIQHGSFATASCLICKYVVDCEAVAGDIFNOVVPKPCPCP-----	176		
Db	314	NIDTLEQVAGIORIVTECHGGSFSTACTCKRFKCMADALADTFAORIPVPCPOCPKPEOS	373		
Qy	177	-----ADEPL-----AAMKPEIYFEGENLPEQGFHRAKMDKQKDEVDLLIYGSSLKVR	223		
Db	374	VDASAAYVTEEBELROLVENGIMKPDIVFEGEGLPDEHYHMAIDKDYCDLLIYIGSSLKVR	433		
Qy	224	PVALIPSSIPHEVPOILINRE 244			
Db	434	PVALIPSSIPATVPQILINRE 454			

```

: RESULT 9
: US-60-191-637-10901
: Sequence, 10901, Application US/60191637
: GENERAL INFORMATION:
: APPLICANT: Ventler, J. Craig
: TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
: TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: CL000392
: CURRENT APPLICATION NUMBER: US/60/191,637
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 42660

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10901
;
; LENGTH: 823
;
; TYPE: PRY
;
; ORGANISM: DROSOPHILA
US-60-191-637-10901

```

Query Match	66.6%	Score 883.5	DB 22	Length 823
Best Local Similarity	63.6%	Pred. No. 9.9e-90		
Matches 166, Conservative	38	Mismatches 38	Indels 19	Gaps 2

Qy	3	NILSPPKRRKKRINTIEDZAVKLLQOEKRTIVLTGAGVSCSIPDFRSRDGIYARLAV	62
Db	194	HLNLPKRNRKLASVNFDDVLSLYKKSQKITVLTLGAGVSSCSIPDFRSNIGIYARLAV	2533
Qy	63	DEPDLDPQAMDIETFKRDPPEPFKAKEIYPCQOFQPSLCHKFTALSDEKGLLRNYTQ	1222
Db	254	DEPDLDPQAMFINDIYFKRDRPRPYKFAREIYPCGFQSPCHRRIFKMLETKGLLRNYTQ	3131
Qy	123	NIDTLEQVAGIORILQCGSFATFASCLICXKVCCEAVRGQINQVQVPRCPDPC-----	176
Db	314	NIDTLEKAVAGIORVTECHGSSSTASCTKCRKCANADARADIFQRIPLVCPQCPKNEQS	3737
Qy	177	-----ADEPL-----AIKRPITVEFGENLPQOFPRAMKYDKDEVDLLIVIGSSLVR	2233
Db	374	VDASAVAVTEEEELROLVENGIMKPKPIVEFGELPDEYHTVATNDRQVODLLIVIGSSSLVR	4333
Qy	224	PVALPSSIPHEPQIILNRE	244
Db	434	PVALPSSIPATVPOIILNRE	454

```

RESULT 10
US-60-191-681-8549
: Sequence 8549, Application US/60191681
: GENERAL INFORMATION:
: APPLICANT: Li, Peter, W.D.
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
: TITLE OF INVENTION: USES THEREOF.
: FILE REFERENCE: C1000390
: CURRENT APPLICATION NUMBER: US/60/191,681
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 30973
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8549
: LENGTH: 823
: TYPE: PRT
: ORGANISM: DROSOPHILA
US-60-191-681-8549

```

```

Query Match          68.66; Score 883.5; DB 22; Length 823;
Best Local Similarity 53.66; Prd. No. 9; ge-90;
Matches 166; Conservative 38; Mismatches 38; Indels 19; Gaps

OY      3  NILSEPRKKRRKRDINTEDAVKLLQECKKIIVLTGAGVSVCSGIPDFRSRDGIYARLAV 62
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      194  HLLNEPKRRNKLASVNEFDVVISLTKKRSOKIIVLTGAGVSVCSGIPDFRSRNGIYARLAAH 253
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY      63  DEPDLPDQAMFDELYEPFKDPRPEFKAKETIYPCQOPSLCHKRIKALSDSEKGLRNAYTO 122
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      254  DFPDLPDQAMFDINFRKDRPRPKARELEIYEPQFSPSCHEKRIKLETKGLRNAYTO 313
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY      123  NIDTLAEQVAGIQRILLOCHGSFATASCLICKKVCACDEVRGDIFNQVVRPCRP----- 176
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      314  NIDTLERVAGIQRVIECHGSFSTASCCKRCKCANADLRADIDFAQRIPVPCOPNKEGS 373
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY      177  -----ADDEPL-----AIKKPIIVFGEMLPEQFRRAKRYKQDEVDLLIIVGSSLSKVR 223
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      374  VDAVAAYVTEELRLQVENGINKRPDIIVEGELPRDEHYIYNATNDYDCDLLIIVGSSLSKVR 433
          :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```
QY      224 PVALIPSSIPHEVPQILNRE 244
      ||| ||||| |||||
Db      434 PVAHIPSSIPATVPQILNRE 454
```

```

RESULT 11
US-09-270-767-44940
; Sequence 44940, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44940
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
; US-09-270-767-44940

```

Query Match	58.0%;	Score 747;	DB 16;	Length 286;
Best Local Similarity	54.4%;	Pred. No. 4.2e-75;		
Matches 148;	Conservative 27;	Mismatches 29;	Indels 68;	Gaps 3

Qy	41	VSVSGIDPFGSRGIYARLAVDPDPDPPDAMPDIEFRKDPFFKFAKEIYGGOF	100
Db	1	VVSGIGDPFGSTNGIYARLAVHDPDLPDPQAMPDINFKRDPPIYKFAEIRYGEFQ	60
Qy	101	SLCHKEFLALSDKEGKLLRNNTQNTIDTLEQVAGIQRILICQHSSEATASCLICKYKDCAV	160
Db	61	SPCHKEFIMETKRLKRLNRYQNTIDTLERVAGIQRIVIECHSEFSTASCICRFKCNADAL	120
Qy	161	RGDFENOVVPCPCPCP-----ADEPL-----AIMSEPLVPRGE-----	199
Db	121	RADIFAQKRIYPCFQCQPKKESVDASAAYVEEELRQLYENGIMKPDIVYFGEQCHMYTD	180
Qy	194	-----NPEQFHRAMKYDKDEVDL	212
Db	181	LLYIINSLGKINARISNSTIKRSYVLRDKLKLKIQSNAIYHIGLDEYHTWATDKDVCDL	240
Qy	213	LIVIGSSLKVRPALIPSSIPHEYPQIILNNG	244
Db	241	LIVIGSSLKVRPAHIPSIPATVPQIILNNE	272

```

RESULT 12
US-09-461-580A-9
: Sequence 9, Application US/09461580A
: GENERAL INFORMATION:
: APPLICANT: Guarente, Leonard
: APPLICANT: Imai, Shin-Ichiro
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 128
: TYPE: PRT
: ORGANISM: Mus musculus
: US-09-461-580A-9

```

Query Match	53.0%; Score 682; DB 18; Length 128;
-------------	--------------------------------------

Best Local Similarity: 100.0%; Pred. No. 2.5e-68;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	34	IVLTGAGVAGVSCGIPDFEFSRDGIYAARLAVDEPDLDPDQAMDIETFKRDPFFPKFAKEI	93
Db	1	IVLTGAGVAGVSCGIPDFEFSRDGIYAARLAVDEPDLDPDQAMDIETFKRDPFFPKFAKEI	60
QY	94	YPGGQPSLCHKFTALSPKESGKLLRNTQNTIDTLEQVAGIORTIQCHGSFPTASCLICKY	155
Db	61	YPGGQPSLCHKFTALSPKESGKLLRNTQNTIDTLEQVAGIORTIQCHGSFPTASCLICKY	128
QY	154	KVDCGAAYR	161
Db	121	KVDCGAAYR	128

```

RESULT 13
US-09-248-796-19599
: Sequence 19599, Application US/09248796
: GENERAL INFORMATION:
: APPLICANT: Keith Weinstock et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBII
: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.132
: CURRENT APPLICATION NUMBER: US/09/248,796
: CURRENT FILING DATE: 1999-02-12
: NUMBER OF SEQ ID NOS: 28206
: SEQ ID NO 19599
: LENGTH: 339
: TYPE: PR1
: ORGANISM: Candida albicans
US-09-248-796-19599

```

Query Match	45.88;	Score 589.5;	DB 16;	Length 339;
Best Local Similarity	44.68;	Pred. No. 2.7e-57;		
Matches 123;	Conservative 37;	Mismatches 73;	Indels 43;	Gaps 4

Qy	.71	QAMDEIEYFRKDPRPEFKAEIYPGOFPSLCKHFKIALDSKESKLLRLNTQNTIDLEQV	130
Db	71	QEVFDLDFLNDPMIFYSIAIMHILPPNHYISPLHSFLTKLDDKNTLLRNTQNTIDNLESY	133
Qy	131	AGI--ORLQCHGSFATASCILCKYKDCVCEVRDIFNOVYPRCPRC-----	175
Db	131	AGIKKENLQCHGSFATASCILCTGKYKDGELIIFPEIKELKEIYPRCKNEVQSLKGGKK	190
Qy	176	-----PAD-----EPALIMKEIYFEGENLPEQFHRAMKYDKE	205
Db	191	TKSKSKKKKKKKRPPYDDEEEEGEYFHEGSFGYMKPDITFFGQLPENFKIAINODINK	250
Qy	210	VDLLITIGSSILKYRPAVLAIPSSIHVEYQQLINREP	245
Db	251	VDLVAIVTSTSLKVAVPADVIGKIPENHPIQLINLNDP	286

```

RESULT 14
US-60-096-409-19599
: Sequence 19599, Application US/60096409A
:
: GENERAL INFORMATION:
:
: APPLICANT: Kelin Weinstock et al
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
:
: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCE: GTC-016P
:
: CURRENT APPLICATION NUMBER: US//60/096,409A
:
: CURRENT FILING DATE: 1998-08-13
:
: NUMBER OF SEQ ID NOS: 28206
:
: SEQ ID NO 19599
:
: LENGTH: 339
:
: TYPE: PRT

```

: ORGANISM: Candida albicans
US-60-096-409-19599

Query Match 45.8%; Score 589.5; DB 22; Length 339;
Best Local Similarity 44.6%; Pred. NO. 2.7e-57;
Matches 123; Conservative 37; Mismatches 73; Indels 43; Gaps 4;

QY 11 RRRKRDINTIEDAVKLLQECRIIVLTGAGVSVSCGIPDFRSRDIYARLAVDPDLPDP 70
DB 13 RRRREDITTINALKLIENSKNIMVTGAGISTGIDPFRSSGCFYS--MIGHGLSDP 70
QY 71 QAMEDIETFRKDPPEPFKAKIYPOPSLCHKFIALSDKEGKLLRNTONTIDLEQV 130
DB 71 QEFVLDLFLNDPNIFYSAIHMILPPNHIYSPLSFIKLQDKNFLRNTONTIDNLESY 130
QY 131 AGI--QRILQCHGFATASCLCKKVKDCEAVRGDIFENVVPRCPRC----- 175
DB 131 AGIHKEMLIQCHGFATASCTICGYKVDGEIIEPIKKEIIPYCPKNEVKOSILKGRK 190
QY 176 -----PAD-----EPLAIKPEIYFPGENLPEOFHRAKRYKDE 209
DB 191 TKSRSKKKKKKKKRKYDDEEEBEGEYFHESRGVAKPDITFFGEQDPENFKAIINODINK 250
QY 210 VDLIVIGSLKVPVALIPSSIPEHVPOILINREP 245
DB 251 VDLVIVIGSLKVPVALIPSSIPEHVPOILINREP 286

RESULT 15

US-09-461-580A-19
: Sequence 19, Application US/09461580A
: GENERAL INFORMATION:
: APPLICANT: Guarente, Leonard
: APPLICANT: Imai, Shin-ichiro
: APPLICANT: Armstrong, Christopher
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FaastSeq for Windows Version 4.0
: SEQ ID NO 19
: LENGTH: 106
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-461-580A-19

Query Match 43.2%; Score 556; DB 18; Length 106;
Best Local Similarity 100.0%; Pred. NO. 2.7e-54;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 IIVLTGAGVSVSCGIPDFRSRDIYARLAVDPDLPDPQAMEDIETFRKDPPEPFKAKE 92
DB 1 IIVLTGAGVSVSCGIPDFRSRDIYARLAVDPDLPDPQAMEDIETFRKDPPEPFKAKE 60
QY 93 IYPOFOPSLCHKFIALSDKEGKLLRNTONTIDLEQVAGIORILQ 138
DB 61 IYPOFOPSLCHKFIALSDKEGKLLRNTONTIDLEQVAGIORILQ 106

Search completed: February 16, 2001, 10:49:14
Job time: 455 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:43:16 ; Search time 76.21 Seconds
(without alignments)
57.431 Million cell updates/sec

Title: US-09-461-580A-9

Perfect score: 682
Sequence: 1 IVLTGAGVSVSCGIPDFRSR.....SFATASCLCKYKVDCEAVR 128

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.36.*
1: /SIDSL/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	326.5	47.9	212	20	V48540	Human breast tumor
2	304.5	44.6	257	21	V73360	HTK clone 2495790
3	84	12.3	205	19	W98718	H. pylori GHP0 808
4	72	10.6	956	14	R32356	Excitatory amino aci
5	70.5	10.3	245	20	V14064	S. pneumoniae Orn1
6	70.5	10.3	245	20	V03651	S. pneumoniae resp
7	70.5	10.3	245	20	W69436	Streptococcus pneu
8	70.5	10.3	245	20	W89433	Streptococcus pneu
9	68.5	10.0	413	20	V07057	Renal cancer assoc
10	64.5	9.5	72	20	V19564	Amino acid sequenc
11	64.5	9.5	110	20	V27041	Amino acid sequenc
12	64.5	9.5	696	20	V13359	Amino acid sequenc

13	64.5	9.5	696	21	V70671	Human PRO266 prote
14	64.5	9.5	720	17	R92227	Thermostable enzym
15	64	9.4	746	17	W05260	Chromatin regulato
16	63.5	9.3	441	16	R73049	Human acetyl hydro
17	63.5	9.3	755	20	W81365	Human prothormone c
18	63.5	9.3	2408	13	R24306	Translation of ORF
19	62.5	9.2	360	20	V06245	Human FAK-related
20	62.5	9.2	441	16	R71924	Human acetyl hydro
21	62.5	9.2	1052	17	R88576	Human focal adhesi
22	62	9.1	985	21	V70073	Secreted protein e
23	61.5	9.0	452	21	W85465	Human growth-assoc
24	61.5	9.0	452	21	V57560	Human growth-assoc
25	61.5	9.0	621	21	V57559	Human growth-assoc
26	61.5	9.0	680	17	R90618	Sulfolobus acidoca
27	61.5	9.0	945	20	V35612	C. pneumoniae prot
28	61.5	9.0	1024	18	W19604	Mycoplasma genital
29	61	8.9	425	19	W61487	Chicken sonic hedg
30	61	8.9	619	14	R42090	c-rel protein. Ho
31	61	8.9	721	18	W34454	Racilius subtilis
32	61	8.9	746	18	W34455	Racilius subtilis
33	60.5	8.9	296	20	V37329	Protein involved i
34	60.5	8.9	441	16	R71923	Human acetyl hydro
35	60	8.8	410	20	W97090	A laif enzyme of S
36	60	8.8	709	21	V81802	Prostaglandin tran
37	60	8.8	709	21	V77455	Human dendritic ce
38	60	8.8	2209	3	P20037	Sequence encoded b
39	59.5	8.7	260	21	V81678	Streptococcus pneu
40	59.5	8.7	274	20	V43931	Fission yeast prot
41	59.5	8.7	282	20	V37026	Chlamydia trachoma
42	59.5	8.7	297	20	W95686	Schizosaccharomyce
43	59.5	8.7	398	20	V34810	Chlamydia pneumoni
44	59.5	8.7	441	16	R71920	Human acetyl hydro
45	59.5	8.7	441	16	R71921	Human acetyl hydro

ALIGNMENTS

RESULT 1	
Y48540	Y48540 standard; Protein: 212 AA.
XX	
AC	Y48540;
XX	
DT	08-DEC-1999 (first entry)
XX	
DE	Human breast tumour-associated protein 1.
XX	
KW	Expressed sequence tag; EST: human; breast; cancer; gene therapy;
KW	treatment; tumour; cytostatic; medicament.
XX	
OS	Homo sapiens.
XX	
PN	DE19813839-A1.
XX	
PD	23-SEP-1999.
XX	
PF	20-MAR-1998; 98DE-1013839.
XX	
PR	20-MAR-1998; 98DE-1013839.
XX	
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.
XX	
PI	Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;
DR	WPI: 1999-528981/45.
XX	
DR	N-PDSB; Z33613.
XX	
PT	Human nucleic acid sequences and protein products from tumor breast
PT	tissue, useful for breast cancer therapy -
XX	
PS	Claim 25; 143; 188pp; German.
XX	

SQ Sequence 205 AA;

OY 36 DPQAMEDIIEFRKDRPREFEKFAKEIYPGQVPSLCHFIALSREKGLLRNTQNIDTLE 95
:||:|:::||:||||:
DB 21 npqkvld-fynqrriqlf----evypnk-----ahkalaelaekhyg-vnilitgnvdllh 68

```
QY      96 QVAGIQRILOCHG 108
      :  ||  |||  ||
Db      69 eragssrllhnhg 81
```

RESULT	4
R32356	
ID	R32356 standard; Protein; 956 AA

DT 21-JUN-1993 (first entry)

Excitatory amino acid receptor 1.

KW EAAR-1; brain; kainate; glutamate; assay; CNS

FH	Key	Location/Qualifiers

FT /note= "signal peptide"

FT /note= "mature protein"

PN EP529994-A.

PD 03-MAR-1993.

PF 25-AUG-1992; 92EP-0307723.

PR 26-AUG-1991; 91US-0750090.

PA (ALLE-) ALLELIX BIOPHARMACEUTICALS INC

PI Kamboj R, Nutt SL, Shekter L, Wosnick MA;

DR WPI; 1993-069001/09.

DR N-PSDB; Q36915.

PT Polynucleotide sequence encoding human excitatory amino acid-1

PT screening for new CNS drugs

PS Disclosure; Fig 1; 28pp; English.

CC Human hippocampal cDNA was subjected to PCR amplification using
CC primers corresp. to regions of the rat Glut1 gene. The amplified
CC prod. was labelled and used to probe a cDNA library from the same
CC source and 50 positive clones isolated; one contained a 1058 bp
CC sequence. This was labelled and used to screen a different

hippocampal cDNA library to identify two overlapping clones from which the 67 kb phagemid pBS/HumEppria (ATCC 75063) contg. the entire open reading frame for EAAR1 receptor was constructed. This complete insert could be removed as a 3.7 kb NotI fragment. The sequence encodes the excitatory amino acid -1 receptor (R) or a kainate binding fragment of it. (R) binds glutamate and has ligand binding properties characteristic of kainate-type EAAR1 receptors. Cells contg. the gene may be used to assay cpds. for possible use in treatment of CNS disorders.

SQ Sequence 956 AA;

Query Match	10.6%	Score 72	DB 14	Length 956
Best Local Similarity	28.9%	Pred. NO. 4.9		
Matches 24	Conservative 9	Mismatches 24	Indels 26	Gaps 5

```

QY      42 DIEFRRKDPREEFKRAKEIYPQGFQPSLCHKFTLSDKEKLLRWYQNIITLEQVAGIO 101
      ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     115 evphfkvaapefvkf-----qfq-----rftln-----lhpnsctlsavagil 154

```

```

QY      102 RIQCHGSFATASCLICKYKVC 124
          | : ||| | : |
Db      155 nfn-----ttaclica-kaec 171

```

RESULT . 5
Y14064
ID Y14064 standard; Protein; 245 AA

AC Y14064;

DT 16-JUL-1999 (first entry)

DE S. pneumoniae Ornithine carbamoyltransferase

KW Ornithine carbamoyltransferase; argF/arcB family polypeptide; diagnosis;
KW infection; bacteremia

KW pneumonia; bacteraemia; sinusitis; pleural empyema; endocarditis;

XXXXXXXXXXXX

XX
XX
55013475 20

[illegible][illegible][illegible][illegible]

XX

XX

DR N-PSDB; X57965.

PT New Streptococcus pneumoniae ornithine carbamoyltransferase family

PT antibacterial drugs, and for diagnosis and treatment of Streptococcal

2
1
1
2
2
2
2
1
2
2
1
1
1

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

CC carbamoyltransferase protein is a (argF/argB) family polypeptide. (I) and CC polynucleotides (II) encoding it are useful for diagnosing diseases due to an infection of an organism with the ornithine carbamoyltransferase gene by determining the nucleic acid sequence encoding (I), and/or CC analysing for the presence or amount of (I). They can diagnose the stage CC and type of infection. (I) is also useful for screening for compounds

PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Wallis NG;
 XX
 DR WPI: 1999-037018/04.
 DR N-PSDB; V82065.
 XX
 PT New response regulator polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Streptococcus pneumoniae infections
 XX
 PS Claim 12; Page 9; 40pp; English.
 XX
 CC The present sequence represents a response regulator (RR) protein from
 CC Streptococcus pneumoniae. RR proteins and polynucleotides are useful as
 CC research reagents and as materials for discovery of treatments of, and
 CC for diagnosing diseases, particularly for diagnosing the stage and type
 CC of an infection using PCR. Diseases related to expression or activity of
 CC RR proteins can be diagnosed by determining the nucleotide sequence
 CC encoding RR, or by analysing for the presence or amount of RR proteins.
 CC RR proteins can be used to screen for agonists and antagonists
 CC (antibacterial compounds) which interact with RR proteins by observing
 CC the binding, or stimulation or inhibition of RR protein activity.
 CC Treatment of diseases related to protein activity can be treated by
 CC administering antagonists to prevent expression, and RR proteins to
 CC enhance expression levels of the protein. RR proteins and polynucleotides
 CC can be administered as a vaccine to protect against disease, and can
 CC prevent adhesion of bacteria to matrix proteins, and are useful for use
 CC on wounds and body implants to prevent bacterial infection. Antibodies
 CC are useful for treating bacterial infections, and can isolate or identify
 CC clones expressing RR proteins, and can purify the proteins by affinity
 CC chromatography. Diseases diagnosed, prevented or treated include: otitis
 CC media; conjunctivitis; pneumonia; bacteremia; sinusitis; pleural empyema;
 CC endocarditis and particularly meningitis, and infections caused by
 CC Streptococcus pneumoniae.
 XX
 SQ Sequence 245 AA;

Query Match 10.3%; Score 70.5; DB 20; Length 245;
 Best Local Similarity 24.1%; Pred. No. 1.2;
 Matches 34; Conservative 20; Mismatches 30; Indels 57; Gaps 9;

QY 16 DFRSRDGIYARLAVDPDPQAMF-DIEYFRKDRPFRKFA-----KEIYPGQRPQS 68
 Db 117 efesr-----letallyansqskslaedcfyfs-----kfaqfgyfkeyv----- 159
 QY 69 LCHKFALSDKEGKLBNYQNTDLEOVAGIQRI-----LQCHGSFAT----- 112
 Db 160 -----yletsprhprvll-ykkt-drlleftasleevfkqpsllqchrsfllpanvvhld 213
 QY 113 -----ASCLICKYKV 122
 Db 214 kkekllffpngscllarykv 234

RESULT 8
 W89433
 ID W89433 standard; Protein; 245 AA.
 AC W89433;
 XX
 DT 22-MAR-1999 (first entry)
 XX
 DE Streptococcus pneumoniae response regulator.
 XX
 KM Streptococcus pneumoniae; response regulator; histidine kinase;
 KM infection; antibacterial; otitis media; conjunctivitis; pneumonia;
 KM bacteremia; sinusitis; pleural empyema; endocarditis; meningitis.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN EP885963-A2.

XX
 PD 23-DEC-1998.
 XX
 PF 17-JUN-1998; 98EB-0304797.
 XX
 PR 20-JUN-1997; 97US-0879528.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Wallis NW;
 XX
 DR WPI: 1999-037058/04.
 DR N-PSDB; V82058.
 XX
 PT New response regulator polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Streptococcus pneumoniae infections
 XX
 PS Claim 12; Page 9; 37pp; English.
 XX
 CC The present sequence is a response regulator (RR) from Streptococcus
 CC pneumoniae. RR proteins and polynucleotides are useful as research
 CC reagents and materials for discovery of treatments of and diagnostics
 CC for disease, and particularly for diagnosing the stage and type of an
 CC infection by PCR. Diseases related to expression or activity of RR can
 CC be diagnosed by determining the nucleotide sequence encoding RR, or by
 CC analysing for the presence or amount of RR protein. RR proteins can be
 CC used to screen for agonists and antagonists which interact with RR
 CC proteins by observing the binding, or stimulation or inhibition of RR
 CC protein activity. The agonist can be used in treatment to inhibit RR
 CC activity, and RR proteins can be administered to treat conditions
 CC associated with a lack of RR protein. Diseases can be prevented by
 CC inoculating with RR protein, or a vector which expresses RR protein. RR
 CC proteins and polynucleotides can prevent adhesion of bacteria to matrix
 CC proteins, and are useful for use on wounds and body implants to prevent
 CC bacterial infection. RR antibodies are useful for treating bacterial
 CC infections, and can isolate or identify clones expressing RR proteins,
 CC and can purify the proteins by affinity chromatography. Diseases
 CC diagnosed, prevented or treated include: otitis media; conjunctivitis;
 CC pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis and
 CC particularly meningitis, and infections caused by Streptococcus
 CC pneumoniae.
 XX
 SQ Sequence 245 AA;

Query Match 10.3%; Score 70.5; DB 20; Length 245;
 Best Local Similarity 24.1%; Pred. No. 1.2;
 Matches 34; Conservative 20; Mismatches 30; Indels 57; Gaps 9;

QY 16 DFRSRDGIYARLAVDPDPQAMF-DIEYFRKDRPFRKFA-----KEIYPGQRPQS 68
 Db 117 efesr-----letallyansqskslaedcfyfs-----kfaqfgyfkeyv----- 159
 QY 69 LCHKFALSDKEGKLBNYQNTDLEOVAGIQRI-----LQCHGSFAT----- 112
 Db 160 -----yletsprhprvll-ykkt-drlleftasleevfkqpsllqchrsfllpanvvhld 213
 QY 113 -----ASCLICKYKV 122
 Db 214 kkekllffpngscllarykv 234

RESULT 9
 Y07057
 ID Y07057 standard; Protein; 413 AA.
 AC Y07057;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Renal cancer associated antigen precursor sequence.

XX Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 XX Homo sapiens.
 PN MO9904265-A2.
 XX 28-JAN-1999.
 PD
 XX 15-JUL-1998; 98WO-US14679.
 PE
 XX 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX (LUDWIG INST CANCER RES.
 PA
 XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 DR WPI: 1999-132448/11.
 XX
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT Isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 PS
 XX Disclosure: Page 446-447; 787pp; English.
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods may be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 CC
 XX
 SQ Sequence 413 AA;
 Query Match 10.0%; Score 68.5; DB 20; Length 413;
 Best Local Similarity 28.1%; Pred. No. 4.2;
 Matches 27; Conservative 14; Mismatches 48; Indels 7; Gaps 3;
 QY 1 IVLTGAGYVSCGIPDFSRDGIYARLAVDPDLPDQAMFDIEYFRDPPFFFAKEI 60
 Db 167 ILLeeNaSgcTcTpIdes-ntihKviegqpdpvraGe-ydvpyftKdEdfnsqwl 224
 QY 61 YPGQFQPSL-----CHKFIASDKKGLLRNTYONI 91
 Db 225 ttgqlpyldgfrhKqksaeadvelnvrlaigql 260
 RESULT 10
 ID Y19564 standard; Protein; 72 AA.
 XX
 AC Y19564;
 XX
 DT 14-JUL-1999 (first entry)
 XX

DE Amino acid sequence of a human secreted protein.
 XX
 KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease;
 KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
 KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
 KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
 KW lung disease; thymus disease; digestive disorder; endocrine disorder;
 KW infection; AIDS.
 XX
 OS Homo sapiens.
 XX
 PN WO9922243-A1.
 XX
 PD 06-MAY-1999.
 XX
 XX 23-OCT-1998; 98WO-US22376.
 PE
 XX 24-OCT-1997; 97US-0063387.
 PR 24-OCT-1997; 97US-0062784.
 PR 24-OCT-1997; 97US-0063088.
 PR 24-OCT-1997; 97US-0063089.
 PR 24-OCT-1997; 97US-0063090.
 PR 24-OCT-1997; 97US-0063091.
 PR 24-OCT-1997; 97US-0063092.
 PR 24-OCT-1997; 97US-0063097.
 PR 24-OCT-1997; 97US-0063098.
 PR 24-OCT-1997; 97US-0063099.
 PR 24-OCT-1997; 97US-0063100.
 PR 24-OCT-1997; 97US-0063101.
 PR 24-OCT-1997; 97US-0063109.
 PR 24-OCT-1997; 97US-0063110.
 PR 24-OCT-1997; 97US-0063111.
 PR 24-OCT-1997; 97US-0063148.
 PR 24-OCT-1997; 97US-0063386.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI
 PI Brewer LA, Carter KC, Duan DR, Edner R, Endress GA;
 PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
 PI Kay H, Latleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
 DR WPI: 1999-303069/25.
 DR N-PSDB; X61444.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PT
 XX
 PS Claim 11; Page 449-450; 546pp; English.
 XX
 CC The specification describes cDNA sequences (X61322-X61470) encoding human
 CC secreted proteins (Y19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.
 CC
 XX
 SQ Sequence 72 AA;

[illegible]

	Key	Location/Qualifiers
KM	nephrotoxic; neuroprotective; anticoagulant; immunological disorder;	
KW	lung; pneumonia; skin; psoriasis; kidney; glomerulonephritis; arthritis	
KW	spondyloarthritis; SLE; systemic lupus erythematosus; scleroderma;	
KW	idiopathic inflammatory myopathy; anaemia; thrombocytopenia; diabetes;	
KW	thyroiditis; Grave's disease; demyelinating disease; multiple sclerosis	
KM	Crohn's disease; hepatobiliary disease; hepatitis; asthma; human;	
KM	graft-versus-host-disease.	
XX	Homo sapiens.	
OS		
FT	Modified-site	17..23
FT	Modified-site	/note- "N-myristoylation site"
FT	Modified-site	18..22
FT	Modified-site	/note- "N-glycosylation site"
FT	Modified-site	30..34
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	67..73
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	100..106
FT	Modified-site	/note- "N-myristoylation site"
FT	Modified-site	122..126
FT	Modified-site	/note- "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	180..184
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	222..226
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	253..257
FT	Modified-site	/note-
FT	Modified-site	"N-glycosylation site"
FT	Modified-site	256..260
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	302..308
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	328..334
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	337..348
FT	Binding-site	/note-
FT	Binding-site	"Prokaryotic membrane lipoprotein lipid attachment site"
FT	Modified-site	343..349
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	354..360
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	363..367
FT	Modified-site	/note-
FT	Modified-site	"N-glycosylation site"
FT	Modified-site	366..370
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	416..420
FT	Modified-site	/note-
FT	Modified-site	"N-glycosylation site"
FT	Modified-site	465..471
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	493..499
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	573..577
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	595..599
FT	Modified-site	/note-
FT	Modified-site	"N-glycosylation site"
FT	Modified-site	598..604
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	603..609
FT	Modified-site	/note-
FT	Modified-site	"N-myristoylation site"
FT	Modified-site	608..612
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	646..650
FT	Modified-site	/note-
FT	Modified-site	"cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	655..659
FT	Modified-site	/note-
FT	Modified-site	"N-glycosylation site"
FT	Modified-site	657..661
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	666..670
FT	Modified-site	/note-
FT	Modified-site	"Casein Kinase II phosphorylation site"
FT	Modified-site	693..697

Query Match 46.5%; Score 317; DB 2; Length 503;

Query Match	45.3%	Score 309;	DB 2;	Length 471;
-------------	-------	------------	-------	-------------

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:45:26 ; Search time 80.58 seconds
(without alignments)
107.859 Million cell updates/sec

Title: US-09-461-580A-9
Perfect score: 682
Sequence: 1 IVLTGAGVSVSCGIPDFRSR.....SFATASCLICKYKVDCEAVR 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	465	68.2	607	2	T24172	hypothetical prote
2	321	47.1	670	2	S36616	regulatory protein
3	317	46.5	503	2	S59638	HST1 protein - yea
4	315.5	46.3	331	2	T18247	transcription regu
5	309	45.3	471	2	T39571	probable regulator
6	304	44.6	562	1	RGBYS2	regulatory protein
7	299.5	43.9	357	2	S59678	HST2 protein - yea
8	281.5	41.3	332	2	T40929	transcription regu
9	278.5	40.8	381	2	JC4639	silent information
10	230.5	33.8	246	2	A72370	regulatory protein
11	218.5	32.0	247	2	H69827	conserved hypothet
12	213.5	31.3	447	2	S54631	HST3 protein - yea
13	207.5	30.4	253	2	H69283	transcription regu
14	185	27.1	247	2	D72562	hypothetical prote
15	181	26.5	245	2	C69459	transcription regu
16	175.5	25.7	415	2	T50106	hst4p (imported) -
17	160	23.5	249	2	F71085	hypothetical prote
18	159	23.3	239	2	C70486	conserved hypothet
19	153.5	22.5	370	2	S52699	hypothetical prote
20	152	22.3	250	2	C75101	transcription regu
21	151.5	22.2	287	2	T22324	hypothetical prote
22	149.5	21.9	259	2	E82191	probable nicotinat
23	143.5	21.0	233	2	G81307	probable transfera
24	143.5	21.0	279	2	E64856	hypothetical prote
25	141.5	20.7	287	2	T22325	hypothetical prote
26	138.5	20.3	250	2	A83506	probable cobalamin
27	126.5	18.5	234	2	A71838	hypothetical prote
28	123.5	18.1	256	2	A83495	hypothetical prote
29	122	17.9	246	2	G75570	conserved hypothet

30 115 16.9 1095 2 T25520 hypothetical prote
31 111 16.3 237 2 H70554 hypothetical prote
32 101 14.8 87 2 T34648 probable SIR2 faml
33 93 13.6 190 2 T35951 probable SIR2 faml
34 84 12.3 205 2 C64677 conserved hypothet
35 82 12.0 320 2 E69264 pyruvate formate-1
36 74 10.9 527 2 S75829 hypothetical prote
37 73.5 10.8 197 2 C49247 mezozoite surface
38 73.5 10.8 614 2 T19539 hypothetical prote
39 73.5 10.8 1289 2 S69689 hypothetical prote
40 72 10.6 486 2 B64532 threonine synthase
41 72 10.6 956 2 JH0826 glutamate ionotrop
42 71 10.4 510 1 S43516 carboxypeptidase C
43 70 10.3 627 2 T19542 hypothetical prote
44 69 10.1 176 2 D49247 mezozoite surface
45 69 10.1 482 2 H69392 4-hydroxybutyrate

ALIGNMENTS

RESULT 1

T24172
hypothetical protein R11A8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24172
R:Cummings, P.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19849
A:Accession: T24172
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <WIL>
A:Cross-references: EMBL:Z70310; PIDN:CAA94364.1; GSPDB:GNO0022; CESP:R11A8.4
A:Experimental source: clone R11A8
C:Genetics:
A:Gene: CESP:R11A8.4
A:Map position: 4
A:Introns: 18/3; 68/3; 111/3; 206/3; 500/3; 565/3

Query Match 68.2%; Score 465; DB 2; Length 607;
Best Local Similarity 64.8%; Pred. No. 1.5e-39;
Matches 83; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

Qy 1 IVLTGAGVSVSCGIPDFRSRDIYARLAVDFPDPQAMFIEYRKDPFRPFKAKEI 60
Db 149 LVLTGAGVSVSCGIPDFRSKDIYARLSEFPDLPDPTAMFDIYRENPAPFYNFAREI 208
Qy 61 YPGQPSLCHKFTALSKEGKLLRNNTQNTDLEQVAGIQRILOCHGSFATASCLICKY 120
Db 209 FPGQVPSPVSHRFKELETSGRLRLNNTQNTDLEHQTGIRKRVCHGSEKCTCTCGCQ 268
Qy 121 KVDCEAVR 128
Db 269 KYDGNR 276

RESULT 2

S36616
regulatory protein SIR2 - yeast (Kluyveromyces marxianus var. lactis)
N:Alternate names: silent information regulator 2
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
C:Accession: A56048; S36616
R:Chen, X.J.; Clark-Walker, G.D.
Mol. Cell. Biol. 14, 4501-4508, 1994
A:Title: sir2 mutants of Kluyveromyces lactis are hypersensitive to DNA-targeting dru
A:Reference number: A56048; MUID:94277055
A:Accession: A56048
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA

Qy 119 KYKVDCEAV 127
:
Db 128 EKKTVEDV 136

RESULT 11
H69827

conserved hypothetical protein yhdz - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: H69827

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
t, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033

A:Accession: H69827

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-247 <KUN>

A:Cross-references: GB:299109; GB:AL009126; NID:g26333260; PIDN:CAB12804.1; PID:g26333

A:Experimental source: strain 168

C:Genetics:

A:Gene: yhdz

C:Superfamily: conserved hypothetical protein bll20

```

Query Match          32.0%; Score 218.5; DB 2; Length 247;
Best Local Similarity 39.4%; Pred. No. 7e-15;
Matches 50; Conservative 22; Mismatches 44; Indels 11; Gaps 3;

Qy      1 IVLTGAGVSVCGLPDRSRDGIYARLAVDPDLDPQAMFDIEYFRKDRPPFPKFAKEI 60
        :|||||:| ||||| |:: | ::|| || |: ||:
Db      15 VVLTGAGNSTESGIPDRSAGGIWTE-----DASRMEAM-SLDYFLSYPLRFWPKEKL 67
        :||::|:| ||||| |:: | ::|| || |: ||:
Qy      61 Y----PGOFPSLCHKFTALSDKEGLLRNYTONIDTLEQVAGIQRIQLCHGSFATASCL 116
        :||::|:| ||||| |:: | ::|| || |: ||:
Db      68 FQMKMSGSEFPNEGHLHLLAELEKGOGKVDTQNIDGLHKAGSRHRVIELHGSIQTAACP 127
        :||::|:| ||||| |:: | ::|| || |: ||:
Qy      117 ICKYKVD 123
        |::|
Db      128 ACGARYD 134

RESULT 12
S54631
HST3 protein - yeast (Saccharomycetes cerevisiae)
N:Alternate names: protein O2651; protein YOR025w
C:Species: Saccharomycetes cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C:Accession: S54631; S62109; S66891
R:de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54617
A:Accession: S54631
A:Molecule type: DNA
A:Residues: 1-447 <DEH>
R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D.
submitted to the EMBL data Library, October 1995
A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in
A:Reference number: S62108

```

Best Local Similarity 51.2%; Pred. No. 8.7e-24;
Matches 65; Conservative 21; Mismatches 33; Indels 8; Gaps 4;

QY 1 IVLTGAGVSCGIPDRSRDGIYARLAVDFDLPDPQAMFDIEYFRKDPDPFFKFAKEI 60
Db 156 VLVGAGISTGLDFRSDNGFYARLARH--GLSEPFSEMFIDHTFRENDPEIFYTTFARDL 213
QY 61 YP--GQFQPSLCHKFIALSDKEGKLLRNTQNTIDTLEQVAGI--QRILQCHGSFATASCL 116
Db 214 LPETNHYSPS--HAFIRLEKKKMLSTLFTQNTIDNLEKKYGLSDNKLIQCHGSFATATCI 271
QY 117 ICKYKVD 123
Db 272 KCKHKVD 278

RESULT 6
RGBYS2

regulatory protein SIR2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D2714; protein YDL042c; silent information regulator 2
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C:Accession: S05891; S67575
R:Shore, D.; Squire, M.; Nasmyth, K.A.
EMBO J. 3, 2817-2823, 1984
A:Title: Characterization of two genes required for the position-effect control of yeast
A:Reference number: S05891; MUID:85126876
A:Accession: S05891

A:Molecule type: DNA
A:Residues: 1-562 <SHO>
A:Cross-references: EMBL:X01419; NID:g4469; PIDN:CAA25667.1; PID:g4470
R:Paulin, L.; Saren, A.M.; Laamanen, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67560

A:Accession: S67575
A:Molecule type: DNA
A:Residues: 1-562 <PAU>
A:Cross-references: EMBL:Z74090; NID:g1431026; PIDN:CAA98600.1; PID:g1431027; GSPDB:GN00
A:Experimental source: strain S288C
C:Genetics:

A:Gene: SGD:SIR2; MIPS:YDL042c
A:Cross-references: SGD:S0002200; MIPS:YDL042c
A:Map position: 4L
C:Superfamily: regulatory protein SIR2
C:Keywords: DNA binding; transcription regulation

Query Match 44.6%; Score 304; DB 1; Length 562;
Best Local Similarity 48.1%; Pred. No. 3.4e-23;
Matches 62; Conservative 20; Mismatches 43; Indels 4; Gaps 2;

QY 1 IVLTGAGVSCGIPDRSRDGIYARLAVDFDLPDPQAMFDIEYFRKDPDPFFKFAKEI 60
Db 258 LVLTGAGVSLGIPDRSSEGSYSK--IKHLGLDDPQOVFNIFMHPDPSVFYNIANMV 315
QY 61 YPCQFQPSLCHKFIALSDKEGKLLRNTQNTIDTLEQVAGI--QRILQCHGSFATASCLIC 118
Db 316 LPPEKTYSPHSPFKMLQMKGLLRNTQNTIDNLESYAGISTDKLVQCHGSFATATCVTC 375
QY 119 KYKVDCEAV 127
Db 376 HNNLPGERI 384

RESULT 7
S59678

HST2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein LPA2c; protein YPL015c
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C:Accession: S59678
R:Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa
submitted to the EMBL Data Library, August 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A:Reference number: S59677
A:Accession: S59678

A:Molecule type: DNA
A:Residues: 1-357 <HAL>

A:Cross-references: EMBL:U33335; NID:g965076; PIDN:AA68090.1; PID:g965078; MIPS:YPL0
C:Genetics:
A:Gene: SGD:HST2
A:Cross-references: SGD:S0005936; MIPS:YPL015c
A:Map position: 16L
C:Keywords: transmembrane protein
F:218-234/Domain: transmembrane #status predicted <TM>

Query Match 43.9%; Score 299.5; DB 2; Length 357;
Best Local Similarity 50.4%; Pred. No. 5.9e-23;
Matches 61; Conservative 20; Mismatches 35; Indels 5; Gaps 3;

QY 1 IVLTGAGVSCGIPDRSRDGIYARLAVDFDLPDPQAMFDIEYFRKDPDPFFKFAKEI 59
Db 28 IFWVGAGISTGIPDRSPGCTGLYNLA--RLKLYPEAVFDVDFQSDPLPFFYTLAKE 85
QY 60 YPCQFQPSLCHKFIALSDKEGKLLRNTQNTIDTLEQVAGI--RILQCHGSFATASCLIC 117
Db 86 LYPGNFRPSKFHYLLKLFQDKDVLKRVYTONIDTLERQAGVKDILLIEAHGSFAHCICIG 145
QY 118 C 118
Db 146 C 146

RESULT 8
T40929

transcription regulator SIR2 homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C:Accession: T40929

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Aert, R.; Volckaert, G.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21958
A:Accession: T40929

A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-332 <LYN>

A:Cross-references: EMBL:AL21807; PIDN:CAB58129.1; GSPDB:GN00068; SPDB:SPCC132.02
A:Experimental source: strain 972h; cosmid c132
C:Genetics:

A:Gene: SPDB:SPCC132.02
A:Map position: 3
A:Introns: 27/1; 152/3

Query Match 41.3%; Score 281.5; DB 2; Length 332;
Best Local Similarity 44.6%; Pred. No. 3.7e-21;
Matches 58; Conservative 28; Mismatches 39; Indels 5; Gaps 3;

QY 2 VLTGAGVSCGIPDRSRDGIYARLAVDFDLPDPQAMFDIEYFRKDPDPFFKFAKEI 60
Db 32 VMVGAGISTAAGIPDRSPETGIYNNL--QRFNLYAEAVFDLSYFRKNRPFEYLAHEL 89
QY 61 YPCQFQPSLCHKFIALSDKEGKLLRNTQNTIDTLEQVAGI--QRILQCHGSFATASCLIC 118
Db 90 MPEKYRPTTYTHYFIRLLHDKRLQKCYTONIDTLERLAGVPDKALIEAHGSFOYSRCIEC 149
QY 119 KYKVDCEAVR 128
Db 150 YEMATEYVR 159

RESULT 9
J94639

silent information regulator 2-related protein - Leishmania major
N:Alternate names: silent information regulator 2 protein homolog

Db 68 MEKVF--NAOPNKAHQAFELERLGLVKCLITQNVDDLHERAGSRNVHLHGSLRVVRC 125
Qy 117 ICKYKVDCEA 126
Db 126 SCNNSFEVES 135

Search completed: February 16, 2001, 10:45:27
Job time: 228 sec


```

PF 02-MAY-1996; 96WO-EP01818.
XX
PR 10-MAY-1995; 95DE-4016776.
XX
PA (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
XX
PI Jenuwein T, Laible G;
XX
DR WPI; 1996-518672/51.
DR N-PSDB; T43624.
XX
PT New DNA encoding chromatin regulator protein with SET domain - and
PT related vectors, transformed cells, proteins and antibodies, for
PT diagnosis and treatment of cancer
XX
PS Claim 12; Fig 7; 38pp; German.
XX
CC The DNA was isolated by screening a human B cell cDNA library with
CC mixed Drosophila DNA probes based on the conserved SET domains
CC in E(z) and Su(var)3-9. The DNA, and its products, are useful
CC in therapy (esp. gene therapy) and diagnosis of human diseases
CC that involve deregulated chromatin-regulator genes having a SET
CC domain, esp. cancer.
XX
SQ Sequence 746 AA;

Query Match 9.4%; Score 64; DB 17; Length 746;
Best Local Similarity 29.5%; Pred. No. 33;
Matches 33; Conservative 13; Mismatches 46; Indels 20; Gaps 8;

QY 30 DFPDLPDP---QAMFDIEYFRKD-----PRPF-----FKFAKEIYPGQ-FQPSLCHKFI 74
Db 186 dddgdgdpereekqkdlhrddkesrpprkfpkskifeaismfpgktaeelkeyk 245
QY 75 ALSDKE--GKLLRNYTONIDTLEQVAGIQRIQLQCHGSFATASCLICKYKVD 124
Db 246 elteqgipgalppectpnldg-pnaksvqreqslh-sfhtlfcrrc-fkydc 294

```

Search completed: February 16, 2001, 10:43:17
Job time: 99 sec

61 YPGQFQPSLCHKFIALSDKEGKLLRNYQNIPILEQVAGI--QRILQCHGSFAIASCLIC 118

DB 372 LPPEMYSPLSFITMIDKDKLNNYNTQNDLNESTYAGVEPEKNVQCHGSPATASCVC 431
 QY 119 KYKDCAV 127
 DB 432 HMKIGERI 440

RESULT 2
 HST1_YEAST STANDARD: PRT: 503 AA.

AC P53685;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST1 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 1).
 GN HST1 OR YOL068C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPH1;
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 in silencing, cell cycle progression, and chromosome stability.";
 RL Science Dev. 9:2888-2902(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GRY 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97321807; PubMed-9178509;
 RA Tzerima M., Katsoulou C., Alexandrak D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast
 chromosome XV reveals eight known genes and ten new open reading
 frames including homologues of ABC transporters, inositol
 phosphatases and human expressed sequence tags.";
 RL Yeast 13:583-589(1997).
 CC -1- FUNCTION: INVOLVED IN TELOMERIC SILENCING, AS OVEREXPRESSION
 CC RESTORES SILENCING AT HMR IN SIR2 MUTANTS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U39041; AAA81033.1; -;
 DR EMBL: L47120; AAB38430.1; -;
 DR EMBL: Z74810; CAAG9078.1; -;
 DR SCD: S0005429; HST1.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN_FING 318 345 C4-TYPE (POTENTIAL).
 FT ZN_FING 318 345 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 503 AA; 57702 MW; 4CDF2799E4135ABB CRC64;

Query Match 46.5%; Score 317; DB 1; Length 503;
 Best Local Similarity 48.8%; Pred. No. 1,4e-26;
 Matches 63; Conservative 23; Mismatches 39; Indels 4; Gaps 2;

OY 1 IVLTGAGVSCGIPDFSRSDGIYARLAVDPDPDPOAMFDIEYFRKDPFFKFAKEI 60

DB 204 LVLTGAGVSTSGIPDFSRSSSEGFYSK--IRHLGLEDDPODVFNLDIFLDDPSVFVIAHAMV 261
 QY 61 YPGGFOPSLCHKHFTALSDKEGKLLNNYNTQNDLEOVAGI--QRILQCHGSPATASCIC 118
 DB 262 LPPEMYSPLSFITMIDKDKLNNYNTQNDLNESTYAGIDPDKLVQCHGSPATASCVC 321
 QY 119 KYKDCAV 127
 DB 322 HMKIGERI 330

RESULT 3
 SIR2_YEAST STANDARD: PRT: 562 AA.

AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 GN SIR2 OR MAR1 OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85126876; PubMed-6098447;
 RA Shore D., Squire M., Nasmyth K.A.;
 RT "Characterization of two genes required for the position-effect
 control of yeast mating-type genes.";
 RL EMBO J. 3:2817-2823(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Paulin L., Saren A.M., Laamanen P.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DACETYLATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X01419; CAA25667.1; -;
 DR EMBL: Z71781; CAA96447.1; -;
 DR EMBL: Z74090; CAA98600.1; -;
 DR PIR: S05891; RCBYS2.
 DR SCD: S0002200; SIR2.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN_FING 254 498 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 372 399 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 562 AA; 63261 MW; 52E6937533654586 CRC64;

Query Match 44.6%; Score 304; DB 1; Length 562;
 Best Local Similarity 48.1%; Pred. No. 4.1e-25;
 Matches 62; Conservative 20; Mismatches 43; Indels 4; Gaps 2;

OY 1 IVLTGAGVSCGIPDFSRSDGIYARLAVDPDPDPOAMFDIEYFRKDPFFKFAKEI 60
 DB 258 LVLTGAGVSTSGIPDFSRSSSEGFYSK--IKHLGLDDPODVFNNTIEMHDSVFVIANMV 315
 OY 61 YPGGFOPSLCHKHFTALSDKEGKLLNNYNTQNDLEOVAGI--QRILQCHGSPATASCIC 118

Db 316 LPPEKITYSPHSHFKMLQMGKLLRNTONTIDMLSEYAGISTDKLVQCHGSFATACVTC 375
 QY 119 KYKVDCEAV 127
 Db 376 HNNLPGERI 384

RESULT 4

HST2_YEAST
 ID HST2_YEAST STANDARD: PRT; 357 AA.
 AC P53686;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE HST2 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 2).
 GN HST2 OR YPL015C OR LPA2C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPHL;
 RA MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RA Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hall J., Ahmed A., Bussey H., Fortin N., Friesen J.D., Storms R.K.,
 RA Vo D.H., Wang Y., Winnett E.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD FUNCTION IN TELOMERIC SILENCING, CELL CYCLE
 CC PROGRESSION AND CHROMOSOME STABILITY.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U39063; AAA81035.1; -
 DR EMBL; U33335; AAB68090.1; -
 DR SGD; S0005936; HST2.
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 22 250 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 143 173 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 357 AA; 39979 MW; ED281E5B8241AAD0 CRC64;

Query Match 43.9%; Score 299.5; DB 1; Length 357;
 Best Local Similarity 50.4%; Pred. No. 7.3e-25;
 Matches 61; Conservative 20; Mismatches 35; Indels 5; Gaps 3;

QY 1 IYVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDLPQAMFIEYFRKDRPPEFKAKE 59
 ID SIR2_LEITMA STANDARD: PRT; 381 AA.
 AC Q25337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RF).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP SEQUENCE FROM N.A.

RESULT 5
 ID SIR2_CANAL STANDARD: PRT; 515 AA.
 AC O59923;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE REGULATORY PROTEIN SIR2.
 GN SIR2.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC anamorphic Saccharomycetales; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martin J., Johnson A.D.;
 RT "Sir2 gene from Candida albicans.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF045774; AAC09304.1; -
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 234 453 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 352 379 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 515 AA; 57732 MW; 2E4DB259EFA9251F CRC64;

Query Match 42.2%; Score 288; DB 1; Length 515;
 Best Local Similarity 48.8%; Pred. No. 1.9e-23;
 Matches 59; Conservative 24; Mismatches 34; Indels 4; Gaps 2;

QY 1 IYVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDLPQAMFIEYFRKDRPPEFKAKEI 60
 ID SIR2_LEITMA STANDARD: PRT; 381 AA.
 AC Q25337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RF).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-96186914; PubMed-8635734;
 RA Yahaiaoui B., Tabl A., Ouassil A.;
 RT "A Leishmania major protein with extensive homology to silent
 information regulator 2 of Saccharomyces cerevisiae.";
 RL Gene 169:115-118(1996).
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L40331; AAB06804.1; -
 CC KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 CC Nucleic protein.
 CC FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
 CC FT ZN_FING 148 151 POLY-ALA.
 CC FT ZN_FING 152 179 C4-TYPE (POTENTIAL).
 CC FT DOMAIN 271 278 POLY-SER.
 CC FT DOMAIN 305 310 POLY-SER.
 CC SEQUENCE 381 AA; 41958 MW; 5A311630A5D2365 CRC64;
 SO
 Query Match 40.8%; Score 278.5; DB 1; Length 381;
 Best Local Similarity 48.8%; Pred. No. 1.4e-22;
 Matches 60; Conservative 20; Mismatches 36; Indels 7; Gaps 4;
 OY 1 LTVGAGVSVSGIPDFRSRDIYARLAVDPDLPDQAMFDEYFRRKDRPFKFAKE 59
 Db LVLGAGVSVSGIPDFRSRDIYARLAVDPDLPDQAMFDEYFRRKDRPFKFAKE 59
 Db 35 LVLGAGVSVSGIPDFRSRDIYARLAVDPDLPDQAMFDEYFRRKDRPFKFAKE 92
 OY 60 --IYVGPQPSLCHKRFLALSDKEGKLRLNRYTONIDTLEQVAGI--QRILQCHGSFATASC 115
 Db LNLWPGHQPFAVHHFIRLLDDEGRLLRCCTQNIIDGLEKAGVSELLVEAHGSAALAC 152
 OY 116 LIC 118
 Db 153 IEC 155
 RESULT 7
 HST3_YEAST STANDARD; PRT; 447 AA.
 AC P53687;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST3 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
 GN HST3 OR YOR025W OR OR26.15.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPH1;
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RT Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RX de Haan M., Maarse A.C., Grievell L.A.;
 RT Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
 CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
 CC AND GENOMIC MAINTENANCE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U39062; AAB81034.1; -
 CC DR EMBL; X87331; CAA60741.1; -
 CC DR EMBL; Z74933; CAA99215.1; -
 CC DR SGD; S0005551; HST3.
 CC KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 CC Nucleic protein.
 CC FT DOMAIN 50 314 SIR2-TYPE CORE DOMAIN.
 CC FT ZN_FING 195 223 C4-TYPE (POTENTIAL).
 CC SEQUENCE 447 AA; 50524 MW; 59CDB5BD712B7005 CRC64;
 SO
 Query Match 31.3%; Score 213.5; DB 1; Length 447;
 Best Local Similarity 38.3%; Pred. No. 1.7e-15;
 Matches 54; Conservative 16; Mismatches 46; Indels 25; Gaps 4;
 OY 3 LTVGAGVSVSGIPDFRSRDIYARLAVDPDLPDQAMFDEYFRRKDRPFKFAKE 57
 Db LVLGAGVSVSGIPDFRSRDIYARLAVDPDLPDQAMFDEYFRRKDRPFKFAKE 117
 OY 58 KEIYPG--QGPQSLCHKRFLALSDKEGKLRLNRYTONIDTLEQVAGI----- 101
 Db LNLWPGHQPFAVHHFIRLLDDEGRLLRCCTQNIIDGLEKAGVSELLVEAHGSAALAC 177
 OY 102 ----RILOCHGSFATASC 118
 Db 178 WKNDVYQHLHGDMLTLSTCK 198
 RESULT 8
 HST4_YEAST STANDARD; PRT; 370 AA.
 AC P53688;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 4).
 GN HST4 OR YDR191W OR YD9346.03.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RT Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP GENE NAME.
 RC STRAIN-GRY 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RT Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";

```

-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  -----
DR  EMBL; AE000531; AAD07166.1; -
DR  TIGR; HP0098; -
DR  PFMAM; PF00291; S.T.dehydratase: 1.
DR  PROSITE; PS00165; DEHYDRATASE_SER_THR. 1.
KW  Theonine biosynthesis; Lyase; Pyridoxal phosphate.
FT  BINDING 109 109 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ  SEQUENCE 486 AA; 54706 MW; CD9637799B508988 CRC64;
-----
Query Match 10.6%; Score 72; DB 1; Length 486;
Best Local Similarity 26.0%; Pred. NO. 2.9;
Matches 32; Conservative 16; Mismatches 39; Indels 36; Gaps 6;
OY 23 IVARLAVDPPLDPQAFEDIEYF--RKDPPEFK-----FAKEIYPPQFQPSLCHKFIA 75
Db 56 VFERLGEIPKNLASALKRYENFPNPNPAPFALNRLRYQELHYG--PSLAFKDMA 112
OY 76 LS-----DKEGRLRLRYTONID---TLEQVAGIORILIOCHGSFATASCLICKY 120
Db 113 LQPLASLEPSNLAVGNERYLMLVTSISGDTGPAITESLGMPPVVF-----VVCVLY 161
OY 121 KYD 123
Db 162 PKD 164
-----
RESULT 10
GLK4_HUMAN STANDARD; PRT; 956 AA.
ID GLK4_HUMAN Q16099;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 4 PRECURSOR (GLUTAMATE RECEPTOR
DE KA-1) (KAI) (EXCITATORY AMINO ACID RECEPTOR 1) (EBA1).
GN GRIK4 OR GRIK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC (1)
RN SEQUENCE FROM N.A.
RP TISSUE=HIPPOCAMPUS;
RX MEDLINE=94087185; PubMed=8263508;
RA Ramboj R.K., Schoepf D.D., Nutt S., Shetter L., Korczak B., True R.A.,
RA Rampeard V., Zimmerman D.M., Wosnick M.A.;
RT "Molecular cloning, expression, and pharmacological characterization
RT of huncbA1, a human kainate receptor subunit.";
RL J. Neurochem. 62:1-9(1994).
CC -1- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC MANY SYNAPSES. IN THE CENTRAL NERVOUS SYSTEM, THE POSTSYNAPTIC
CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS.
CC -1- SUBUNIT: FORMS A HETEROMERIC CHANNEL WITH GRIK1 OR GRIK3.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC

```

DR	EMBL: D17586; BAA04510.1; .	
DR	PIR; S43516; S43516.	
DR	HSSP; P08819; 1BCR.	
DR	MEROPS; S10.001; .	
DR	INTERPRO; IPRO01563; .	
DR	PFAM: PF00450; serine_carboxypep. 1.	
DR	PRINTS; PR00724; CARBOXYPEPTASEC.	
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.	
DR	PROSITE; PS00342; MICROBODIES_CTER; 1.	
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.	
KW	Hydrolase; Carboxypeptidase; Glycoprotein; zymogen; signal.	
FT	SIGNAL	1 25
FT	PROPEP	26 36
FT	CHAIN	37 510
FT	ACT_SITE	194 194
FT	ACT_SITE	434 434
FT	ACT_SITE	487 487
FT	CARBOHYD	154 154
FT	CARBOHYD	268 268
FT	CARBOHYD	418 418
FT	SITE	508 510
SEQUENCE	510 AA: 55709 MW: 2889624 DALI37JCF CRC64;	

Query Match	10.48;	Score 71;	DB 1;	Length 510;
Best Local Similarity	22.38;	Pred. No. 4;		
Matches	33;	Conservative	20;	Mismatches 55; Indels 40; Gaps 7,

09 GAGVSVSCIIPEFRSRDGIYARLAVDPDLQPMADIEYFPKDD---RPFFFAKEI 60
| : | : : | :
146 GVQLSYSKNTSDYNMGD--LKTAAD-----SHHFLMKQQLVEYELSNPFY IAGES 194
Db

Db 195 YAGVYVPFLTSHSEVYAGLHDGYKPTINFGKMGVNGVCDFVEDGNALVPFAHGMALISDDI 254

Db 255 YQEAQTACHGNYWNTTDDKCNALYKVD 282

ID	Y376_METJA	STANDARD;	PRT;	728 AA
Y376_METJA				
AC	057821;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	HYPOTHETICAL PROTEIN MJ0376.			
GN	MJ0376.			
OS	Methanococcus jannaschii.			

OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*." ;
RT
RL Science 273:1058-1073 (1996).

CC
----- CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: 067490; AAB8365.1; -
DR TIGR: M0376; -
KW Hypothetical protein.
SQ SEQUENCE 728 AA; 86466 MW; 8BC3DBCA30226B40 CRC64;

Query Match 10.0%; Score 68.5; DB 1; Length 728;
Best Local Similarity 24.2%; Pred. No. 11;
Matches 29; Conservative 21; Mismatches 33; Indels 37; Gaps 7;

QY 18 RSDGDIYARLAVDFPDPQAMPD--IEFRDPRFFFAKFIYGCQPSLCKHFA 75
DB 172 KREGIFDDV-LENPEIYPTLNDPDISFLDYEY-KFQKELYNK-----NKFVM 223
QY 76 L-----SKEGKL-----LRNYTON-----IDTLEOVAGIQRILQCHG 108
DB 224 LFAPCGRKTEGALLNALNALNFKRKITLAMPQVTSNMYDRLVKIFGEENVGLFEG 283

RESULT 13
LEU2_SCHPO STANDARD; PRT; 758 AA.
AC 014289;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 3-ISOPROPYLMALATE DEHYDRATASE (EC 4.2.1.33) (ISOPROPYLMALATE
DE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPM1).
GN SPAC93.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McDougall R., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE -> 2-ISOPROPYLMALATE +
CC H(2O) (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2O) -> 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 299262; CAB16402.1; -
DR INTERPRO: IPR000573; -
DR INTERPRO: IPR001030; -
DR PFAM: PF00694; Aconitase-C; 1.
DR PFAM: PF00330; aconitase; 1.
DR PRINTS: PR00415; ACONITASE.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S.
FT METAL 359 359 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 420 420 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 423 423 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 758 AA; 82782 MW; CE78C36828380E47 CRC64;

Query Match 10.0%; Score 68; DB 1; Length 758;
Best Local Similarity 24.8%; Pred. No. 13;

Matches 28; Conservative 16; Mismatches 47; Indels 22; Gaps 4;
QY 3 LTGAGVSGGIDPFRSDIYARLAVDFPDPQAMPDIEFRDPRFFFAKFIY 62
DB 526 VASVSSGSGAGIKFTYVEGIAA-----PLPMANVDPKIIPKFLTKRTGL 574
QY 63 GGFQPSLCKHFIAS-DKEGKLRLNYTONIDTLEOVAGIQRILQCHGSPFAS 114
DB 575 GGF-----AFELRDADGKEIPDFVLRNPEYRHA-----VLVAHDFGCGS 617

RESULT 14
CBP1_HORVU STANDARD; PRT; 499 AA.
AC 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SERINE CARBOXYPEPTIDASE I PRECURSOR (EC 3.4.16.5) (CARBOXYPEPTIDASE
DE C) (CP-M1).
GN CBP1 OR CXP.1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ALBUKONE;
RA Rocher A., Lok F., Cameron-Mills V., von Wettstein D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 88-499 FROM N.A.
RX MEDLINE=88298749; PubMed=3403516;
RA Doan N.P., Fincher G.B.;
RT "The A- and B-chains of carboxypeptidase I from germinated barley
RT originate from a single precursor polypeptide."
RL J. Biol. Chem. 263:11106-11110(1988).
RN [3]
RP SEQUENCE OF 31-296 AND 352-499.
RA Soerensen S.B., Bredam K., Svendsen I.;
RT Primary structure of carboxypeptidase I from malted barley."
RL Carlsberg Res. Commun. 51:475-485(1986).
CC -1- FUNCTION: MAY BE INVOLVED IN THE DEGRADATION OF SMALL PEPTIDES (2-
CC 5 RESIDUES) OR IN THE DEGRADATION OF STORAGE PROTEINS IN THE
CC EMBryo.
CC -1- CATALYTIC ACTIVITY: RELEASE OF A C-TERMINAL AMINO ACID WITH A
CC BROAD SPECIFICITY.
CC -1- SUBUNIT: CARBOXYPEPTIDASE I IS A DIMER, WHERE EACH MONOMER IS
CC COMPOSED OF TWO CHAINS LINKED BY DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOSPERM.
CC -1- DEVELOPMENTAL STAGE: AFTER ONE DAY OF GERMINATION, MAINLY FOUND IN
CC THE SCUTELLUM OF THE DEVELOPING GRAIN; BARLEY DETECTABLE AFTER
CC FOUR DAYS, AND ABSENT FROM THE MATURE GRAIN. A LOWER LEVEL OF
CC EXPRESSION IS SEEN IN THE ALBUKONE BOTH DURING DEVELOPMENT AND
CC GERMINATION.
CC -1- PTM: THREE DISULFIDE BONDS ARE PRESENT.
CC -1- PTM: THE LINKER PEPTIDE IS ENDOPROTEOLYTICALLY EXCISED DURING
CC ENZYME MATURATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10; ALSO KNOWN AS THE
CC SERINE CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y09603; CAI70816.1; -
DR EMBL: J03897; AAA32940.1; -
DR PIR: A25858; CPBHS.
DR PIR: B25858; B25858.
DR PIR: A29226; A29226.

DR HSSP; P08819; 1MHT.
 DR MEROPS; S10.001; -.
 DR INTERPRO; IPR001563; -.
 DR PFAM; PF00450; serine carboxylase. 1.
 DR PRINTS; PR00724; CARBOXYPTASEC. 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00342; MICROBODIES_CTER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolyase; Carboxypeptidase; Glycoprotein; zymogen; signal.
 FT SIGNAL 1 30
 FT CHAIN 1 296
 FT PROPEP 297 351
 FT CHAIN 352 499
 FT ACT_SITE 188 188
 FT ACT_SITE 423 423
 FT ACT_SITE 476 476
 FT CARBOHYD 148 148
 FT CARBOHYD 262 262
 FT CARBOHYD 407 407
 FT SITE 497 499
 FT CONFLICT 102 102
 SO SEQUENCE 499 AA; 54096 MW; 9C6674B14D9DB9BF CRC64;

Query Match 9.9%; Score 67.5; DB 1; Length 499;
 Best Local Similarity 20.6%; Pred. No. 9.2;
 Matches 28; Conservative 22; Mismatches 49; Indels 37; Gaps 6;

OY 5 GAGVSGCIGIDFRSDIYARLAVDPDLPQAMFIEFRKDP-----RPEFFAKFI 60
 DB 140 GVGLSYSKNSVDYETGD---LKTARD-----SHTFLKWFQLYLPEFLSNPEY-IGAES 188
 OY 61 YPGQFQPSLCHKFI-----ALSDK--EGKLLRNYTQNIDTLEQVA 98
 DB 189 YAGYVPTLSHEVVGICGAKPTINRGYWGNGVCOTIFDGNALVPAFGMGILSDEI 248
 OY 99 GIORLOCHGSFATAS 114
 DB 249 YQOASTSCHGNVWNT 264

RESULT 15
 Y104_MYCTU STANDARD; PRT; 504 AA.
 AC 010898;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 53.5 KDA PROTEIN RVO104.
 GN RVO104 OR MTCY251.23.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE-98295987; Pubmed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z74410; CA98917.1; -.
 DR HSSP; P00514; 1BPK.
 DR TUBERCULIST; RVO104; -.
 DR INTERPRO; IPR000595; -.
 DR PFAM; PF00027; CNMP_binding; 1.
 DR PROSITE; PS00442; CNMP_BINDING_3; UNKNOWN_1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 146 166
 FT TRANSMEM 196 216
 FT TRANSMEM 330 350
 SO SEQUENCE 504 AA; 53450 MW; 3FCF8A42E875009F CRC64;

Query Match 9.9%; Score 67.5; DB 1; Length 504;
 Best Local Similarity 26.5%; Pred. No. 9.3;
 Matches 27; Conservative 11; Mismatches 55; Indels 9; Gaps 2;

OY 1 IVLNGAGVSGCIGIDFRSDIYARLAVDPDLPQAMFIEFRKDP-----RPEFFAKFI 60
 DB 53 VAVIGKCYSTHGVVDAMRADGIYVD---DCSDAYAPHESEFDQYTRHVERFFAESWART 109
 OY 61 YPGQFQPSLCHKFI-----ALSDK--EGKLLRNYTQNIDTLEQVAGIOR 102
 DB 110 TAGRTAR-----VLLDDGSGSLAVAGAMLDASADVIGIEQ 145

Search completed: February 16, 2001, 10:52:30
 Job time: 533 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:44:05 ; Search time 51.38 Seconds
(without alignments)
44.735 Million cell updates/sec

Title: US-09-461-580A-9

Perfect score: 682

Sequence: 1 IVLTGAGVSVSCGIPDRSR.....SFATASCLICKYKDCAVR 128

Scoring table: BIOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/PCVUS.COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	10.6	956	1	US-08-185-232A-2
2	72	10.6	956	1	US-08-416-523-2
3	72	10.6	956	3	US-08-789-478-2
4	70.5	10.3	245	2	US-08-874-138-2
5	67	9.8	579	1	US-08-448-196A-8
6	64	9.4	1100	1	US-08-357-596-11
7	64	9.4	1100	2	US-09-003-289-11
8	64	9.4	1100	4	PCR-US95-16435-11
9	63.5	9.3	755	3	US-09-071-101-2
10	63.5	9.3	755	3	US-09-369-618-2
11	63.5	9.3	755	3	US-09-369-617-2
12	62.5	9.2	1052	3	US-08-863-118-1
13	62.5	9.2	1052	3	US-09-377-310-2
14	61.5	9.0	376	2	US-08-485-449-5
15	61.5	9.0	452	3	US-09-059-522-3
16	61.5	9.0	452	3	US-09-382-027-3
17	61.5	9.0	621	3	US-09-059-522-1
18	61.5	9.0	621	3	US-09-382-027-1
19	61.5	9.0	720	2	US-08-840-236-1
20	61.5	9.0	720	2	US-08-505-448A-1
21	61	8.9	425	3	US-08-946-329A-19
22	61	8.9	425	3	US-08-567-357A-19
23	61	8.9	425	4	PCR-US95-15463-19
24	61	8.9	425	4	PCR-US95-15923-19
25	60	8.8	180	1	US-08-447-010-8
26	59.5	8.7	219	2	US-08-204-656B-2
27	59.5	8.7	219	2	US-08-467-046-6
28	59.5	8.7	274	2	US-07-857-224B-30

29	59.5	8.7	282	1	US-08-318-947A-19	Sequence 19, Appl
30	59.5	8.7	282	2	US-08-795-303-19	Sequence 19, Appl
31	59.5	8.7	297	2	US-08-874-347-22	Sequence 22, Appl
32	59.5	8.7	297	3	US-09-093-522-22	Sequence 22, Appl
33	59.5	8.7	441	1	US-08-470-187-8	Sequence 8, Appl1
34	59.5	8.7	441	1	US-08-318-905-8	Sequence 8, Appl1
35	59.5	8.7	441	1	US-08-483-232-8	Sequence 8, Appl1
36	59.5	8.7	441	1	US-08-483-140-8	Sequence 8, Appl1
37	59.5	8.7	441	2	US-08-485-938A-8	Sequence 8, Appl1
38	59.5	8.7	441	2	US-08-910-041-8	Sequence 8, Appl1
39	59.5	8.7	441	3	US-09-328-474-8	Sequence 8, Appl1
40	59.5	8.7	441	3	US-09-100-546-8	Sequence 8, Appl1
41	59.5	8.7	441	3	US-09-010-715-8	Sequence 8, Appl1
42	59	8.7	3567	2	US-07-642-734C-4	Sequence 4, Appl1
43	59	8.7	3567	3	US-08-439-009A-4	Sequence 4, Appl1
44	58.5	8.6	284	1	US-08-411-777-10	Sequence 10, Appl
45	58.5	8.6	284	3	US-09-057-088-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-185-232A-2
Sequence 2, Application US/08185232A
Patent No. 5576205
GENERAL INFORMATION:
APPLICANT: KAMBOJ, Rajender
APPLICANT: NUTT, Stephen
APPLICANT: SHEKTER, Lee
APPLICANT: WOSNICK, Michael
TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
THE EAAT FAMILY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W. Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,232A
FILING DATE: 24-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,090
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/150 ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 956 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-185-232A-2

Query Match 10.6%; Score 72; DB 1; Length 956;
Best Local Similarity 28.9%; Pred. No. 2.2;
Matches 24; Conservative 9; Mismatches 24; Indels 26; Gaps 5;

STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874,138
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 245 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-874-138-2

Query Match 10.38; Score 70.5; DB 2; Length 245;
Best Local Similarity 24.18; Pred. No. 0.49; Mismatches 9;
Matches 34; Conservative 20; Indels 57; Gaps 9;

QY 16 DFRSDGIYARLAVDPDLDPQAMF-DIEYFRKDPPEFKA-----KEIYGFQFQ 68
Db 117 EFESR-----ETALLVANSQDSKSLADCFYFKS-----KFAQFYRFEKVV----- 159
QY 69 LCHKEFALSKEGKLRLNYQNTDLEOVAGIORI-----LQCHGSFAT----- 112
Db 160 -----YLETSRPHRVIL-YRKT-DRLFTASLEEVFKQEPSELQCHRSFLINPANYVLD 213
QY 113 -----ASCLICKYK 122
Db 214 KKEKLEFPNGSCLARYK 234

RESULT 5
US-08-448-196A-8
Sequence 8, Application US/08448196A
Patent No. 5780594
GENERAL INFORMATION:
APPLICANT: CARTER, DANIEL C.
TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN FRAGMENTS
TITLE OF INVENTION: CONTAINING SPECIFIC BINDING REGIONS OF SERUM ALBUMIN OR
NUMBER OF INVENTIONS: RELATED PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NASA
STREET: MARSHALL SPACE FLIGHT CENTER
CITY: HUNTSVILLE
STATE: ALABAMA
COUNTRY: USA
ZIP: 35812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,196A

FILING DATE: 23-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: BROAD JR., ROBERT L.
REGISTRATION NUMBER: 18,757
REFERENCE/DOCKET NUMBER: XX/MS-28402-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 205-544-0021
TELEFAX: 205-544-0258
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-448-196A-8

Query Match 9.88; Score 67; DB 1; Length 579;
Best Local Similarity 25.28; Pred. No. 4.6;
Matches 34; Conservative 21; Mismatches 50; Indels 30; Gaps 7;

QY 12 CGIDFRSDGIYARLAVDPDLDPQAMF-DIEYFRKDPPEFKA-----EIIYGFQFQ 66
Db 276 CNLP--LLEFYCIVLENDVP-AELSKPTEFTEDPHVCQYAEKRSLEISPMQSO 331
QY 67 --PSLCHKEFALSDEKGLRLN-----YTONIDL-----EOVAGIOR---ILQ 105
Db 332 ETPELSEGFLLQSAKEYESLNNKCCFSNPECKYKDGADRMNKAKEFATLKQNCIDLH 391
QY 106 CHGSFATASCLICKY 120
Db 392 EHGEYLFENELLIRY 406

RESULT 6
US-08-357-598-11
Sequence 11, Application US/08357598
Patent No. 5705625
GENERAL INFORMATION:
APPLICANT: Civid, Curt I.
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1100 amino acids

```

;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-357-598-11

```

Query Match	9.48	Score 64	DB 1	Length 1100
Best Local Similarity	27.98	Pred. No. 28		
Matches 29	Conservative 13	Mismatches 34	Indels 28	Gaps 6

QY 27 LAVDFPDLPRQAMFDI--EYRRK--DRPFPAKAEIIPGF-----QPSLCIKFIALSDK 79
 1 :||| : : ||| | | ||| : : ||| | |
 Db 325 LEAEFPGLPEALSFVALVDGYRFLICDSRH--FCKEVAAPRLLEEAEALCHGPITLDF 382

```

QY      80 EGGLELRYNTQNIIDTLEQVAG-----IQRIQCHGSFATASCL 116
      ||      : ||      ::! | : ||      :! :
Db      383 IHKL-----KAAGSLPGSYILRRSPQDYDSFLLTACV 414

```

RESULT 7
US-09-003-289-11
; Sequence 11, Application US/09003289
; Patent No. 5916792

```

; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
;
; NUMBER OF SEQUENCES: 12
;
; CORRESPONDENCE ADDRESS:

```

ADDRESS: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/003,289

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
FAX: 619/678-5000

```

: INFORMATION FOR SEQ ID NO: 11:
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 1100 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: NO. 5916792 Relevant
:

```

```

; MOLECULE TYPE: protein
US-09-003-289-11

```

Query Match	9.48	Score	64	DB	2	Length	1100		
Best Local Similarity	27.98	Pred.	NO.	28					
Matches	29	Conservative	13	Mismatches	34	Indels	28	Gaps	6

OY 27 L A N D F P D L T P D Q A M E D I - - E Y R K - - D R P F F K A K E I Y P G O F - - Q P S L C K F I A L S D K 79
| : | | : : : | | | : | | : | | |
D b 325 L E A E P G L E A L S F V A L V D G Y F R L I C D S R H F - - F C K E V A P R L L E E A E L C H G P I T L D F A 383

```

QY      80 EGKLLRNFTQNIIDLEQVAG-----IQRLQCHGSFATASCL 116
          ||          : ||          ::| | : ||          ::|
Db      383 IHKL-----KAAGSLPGSYLLRSPQDYDSFLLTACV 414

```

RESULT 8
PCT-US95-16435-11
Sequence 11 Amplification BC/WTUS9516435

1. CANDIDATE: THE CANDIDATE
 2. APPLICANT: The Johns Hopkins University School of Medicine
 3. TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
 4. NUMBER OF SEQUENCES: 12
 5. CORRESPONDENCE ADDRESS:

ADDRESS: FISH & RICHARDSON P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA

```

: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:

```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: 11418 1433 1

REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5098

```

: INFORMATION FOR SEQ ID NO: 11
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 1100 amino acids
:
:   TYPE: amino acid
:
:   STRADEDNESS: not relevant
:
:   TOPOLOGY: linear
:
:   MOLECULE TYPE: protein
:
: PCT-US95-16435-11

```

Query Match	9.4%	Score 64	DB 4	length 1100
Best Local Similarity	27.9%	Pred. No. 28		
Matches 29; Conservative 13; Mismatches 34; Indels 28; Gaps 6				

```
QY 27 L A N D F P D L E D P Q A M E D I - E Y F R K - D P R P F F K A E I Y P G F - - Q P S L C K H F I A L S D K 79
      | : | | | : : | | | | | | | : : | | |
Db 325 L E A F P G L E A L S F V A L V D G Y F R L I C D S R H F - F C K E A V A P R L L E E A E L C H G P I T L D F A 383
```

```

QY      80 ECKLELRNTQNIIDLEQVAG-----IQRIQCHGSFATASCL 116
          ||          : ||          : ||          : ||
Db      383 IHKL-----KAAGSLPGSYILRRSPQDYDSFLLTACV 414

```

RESULT 9
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503

```

; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS

```

```

; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
;

```

```

QY      80 EGKLLRNFTQNIIDLEQVAG-----IQRLQCHGSFATASCL 116
          ||          : ||          ::| | : ||          ::|
Db      383 IHKL-----KAAGSLPGSYLLRSPQDYDSFLLTACV 414

```

RESULT 8
PCT-US95-16435-11
Sequence 11 Amplification BC/WTUS9516435

1. CANDIDATE: THE CANDIDATE
 2. APPLICANT: The Johns Hopkins University School of Medicine
 3. TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
 4. NUMBER OF SEQUENCES: 12
 5. CORRESPONDENCE ADDRESS:

ADDRESS: FISH & RICHARDSON P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA

```

: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:

```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: 11418 1433 1

REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5098

```

: INFORMATION FOR SEQ ID NO: 11
:
: SEQUENCE CHARACTERISTICS:
:
:   LENGTH: 1100 amino acids
:
:   TYPE: amino acid
:
:   STRADEDNESS: not relevant
:
:   TOPOLOGY: linear
:
:   MOLECULE TYPE: protein
:
: PCT-US95-16435-11

```

Query Match	9.4%	Score 64	DB 4	length 1100
Best Local Similarity	27.9%	Pred. No. 28		
Matches 29; Conservative 13; Mismatches 34; Indels 28; Gaps 6				

```
QY 27 L A N D F P D L E D P Q A M E D I - E Y F R K - D P R P F F K A E I Y P G F - - Q P S L C K H F I A L S D K 79
      | : | | | : : | | | | | | | : : | | |
Db 325 L E A F P G L E A L S F V A L V D G Y F R L I C D S R H F - F C K E A V A P R L L E E A E L C H G P I T L D F A 383
```

```

QY      80 ECKLELRNTQNIIDLEQVAG-----IQRIQCHGSFATASCL 116
          ||          : ||          : ||          : ||
Db      383 IHKL-----KAAGSLPGSYILRRSPQDYDSFLLTACV 414

```

RESULT 9
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503

```

; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS

```

```

; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
;

```

```
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,101
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E.
REGISTRATION NUMBER: 31,648
REFERENCE/DOCKET NUMBER: 97-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 755 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-09-071-101-2
```

```
Query Match          9.3%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 19;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

QY 3 LTGAGVSCGIPDFRSRDGIYARLAVDFPDL--PPQAMFDIEYFRKDPPEFFKFAKE 59
DB 147 LSGGQIVVS--VLD-----DGI-----EKDHPDLMANYPDLASYDNDYDPDPQPRYTSKE 196
```

```
RESULT 10
US-09-369-618-2
; Sequence 2, Application US/09369618
; Patent No. 6100041
; GENERAL INFORMATION:
; APPLICANT: Lok, S1
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D2
; CURRENT APPLICATION NUMBER: US/09/369,618
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-369-618-2
```

```
Query Match          9.3%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 19;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

QY 3 LTGAGVSCGIPDFRSRDGIYARLAVDFPDL--PPQAMFDIEYFRKDPPEFFKFAKE 59
DB 147 LSGGQIVVS--VLD-----DGI-----EKDHPDLMANYPDLASYDNDYDPDPQPRYTSKE 196
```

```
RESULT 11
US-09-369-617-2
; Sequence 2, Application US/09369617
; Patent No. 6127162
; GENERAL INFORMATION:
; APPLICANT: Lok, S1
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; FILE REFERENCE: 97-05D1
; CURRENT APPLICATION NUMBER: US/09/369,617
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: US 09/071,101
; EARLIER FILING DATE: 1998-05-01
; EARLIER APPLICATION NUMBER: US 60/044,015
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-369-617-2
```

```
Query Match          9.3%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 19;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

QY 3 LTGAGVSCGIPDFRSRDGIYARLAVDFPDL--PPQAMFDIEYFRKDPPEFFKFAKE 59
DB 147 LSGGQIVVS--VLD-----DGI-----EKDHPDLMANYPDLASYDNDYDPDPQPRYTSKE 196
```

```
RESULT 12
US-08-863-118-1
; Sequence 1, Application US/08863118A
; Patent No. 6015893
; GENERAL INFORMATION:
; APPLICANT: CANCER, William G.
; APPLICANT: LID, Edison T.
; APPLICANT: OWENS, Lewis V.
; TITLE OF INVENTION: Oligonucleoside Compounds and Methods for Inhibiting
; FILE REFERENCE: 5470-90
; CURRENT APPLICATION NUMBER: US/08/863,118A
; CURRENT FILING DATE: 1997-05-27
; EARLIER APPLICATION NUMBER: 08/276,843
; EARLIER FILING DATE: 1994-07-18
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1052
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-863-118-1
```

```
Query Match          9.2%; Score 62.5; DB 3; Length 1052;
Best Local Similarity 24.8%; Pred. No. 40;
Matches 25; Conservative 17; Mismatches 40; Indels 19; Gaps 5;

QY 13 GIPDFRSRDGIYARLAVDFPDLPPQAMFDIEYFRKDPPEFFKFAKEIYFGQFQPS 68
DB 719 GYPSRSESEGY-----PSPQHMVGTNHVQSGYPSHGTAMAGSIYPG--QAS 766

QY 69 LCHKFIASDEKGLNANTYQND--TLEQVAGIQRILQCH 107
DB 767 LLDQTDSMNHRPOEIAM--WQPNVEDSTVLDLRIGIGVALPH 806
```

```
RESULT 13
US-09-377-310-2
```

Sequence 2, Application US/09377310B
Patent No. 613031
GENERAL INFORMATION:
APPLICANT: Monte, Brett P.
TITLE OF INVENTION: Antisense Modulation of Focal Adhesion Kinase
FILE OF INVENTION: Expression
FILE REFERENCE: ISPH-0389
CURRENT APPLICATION NUMBER: US/09/377,310B
CURRENT FILING DATE: 1999-08-19
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1052
TYPE: PRT
ORGANISM: Homo sapiens
US-09-377-310-2

Query Match 9.2%; Score 62.5; DB 3; Length 1052;
Best Local Similarity 24.8%; Pred. No. 40;
Matches 25; Conservative 17; Mismatches 40; Indels 19; Gaps 5;

OY 13 GIPDRSRDGIYARLAVDPDLPQAMFDIEYERKDRP-----FFKAKEIYPCQFOPS 68
Db 719 GYPSRSESGY-----PSPQHVQTNHYGVSGYPSGHGITAAGSIYPS--QAS 766

OY 69 LCHKFIALSDKEGKLRYNTONID--TEQVAGIORILQCH 107
Db 767 LLDQDSWNHRPQETAM-WQPVEDSTVLDLNGIGVLPFH 806

RESULT 14
US-08-485-449-5
Sequence 5, Application US/08485449
Patent No. 5824789
GENERAL INFORMATION:
APPLICANT: VANDENBERG, DAVID
TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,449
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20296-20035.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-485-449-5

Query Match 9.0%; Score 61.5; DB 2; Length 376;
Best Local Similarity 25.2%; Pred. No. 12;
Matches 33; Conservative 14; Mismatches 57; Indels 27; Gaps 6;

OY 7 GVSVSC-----GIPDRSRDGIYARLAVDF-----BDLPDQAMFDIEYER 47
Db 241 GTSGSCQFKTCWRAPERRAGALRLERLAFIDTHNNSGAFQPRRLRSLGELYEE 300

OY 48 KDRPFFKFAKEIYRGQPSLCHKFIALSDKEGKLRYNTONIDTEQVAGIORILQCH 107
Db 301 KSP-DECDERDPTSGPT-RGRACNKTSLDLDCGSLCCGRGHNV--LRQT-----RVERCH 352

OY 108 GSFATASCILIC 118
Db 353 CRFHWCCYVLC 363

RESULT 15
US-09-059-522-3
Sequence 3, Application US/09059522
Patent No. 6001607
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN GROWTH-ASSOCIATED METHYLTRANSFERASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,522
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0495 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LINDNOT08
CLONE: 3053783
US-09-059-522-3

Query Match 9.0%; Score 61.5; DB 3; Length 452;
Best Local Similarity 24.3%; Pred. No. 16;
Matches 33; Conservative 17; Mismatches 25; Indels 61; Gaps 9;

Search completed: February 16, 2001, 10:44:07
Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:49:14 : Search time 220.15 Seconds
(without alignments)
88.467 Million cell updates/sec

Title: US-09-461-580A-9

Perfect score: 682

Sequence: 1 IVLTGAGVSVCIGIPDFRSR.....SEFASCLICKYKVCDAVR 128

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 936741 seqs, 152156132 residues

Total number of hits satisfying chosen parameters: 936741

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Pending_Patents-AA:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	682	100.0	128	US-09-461-580A-9	Sequence 9, Appli
2	682	100.0	245	US-09-461-580A-4	Sequence 4, Appli
3	682	100.0	327	US-09-461-580A-12	Sequence 12, Appli
4	682	100.0	737	US-09-461-580A-1	Sequence 1, Appli
5	682	100.0	737	US-09-461-580A-26	Sequence 26, Appli

6	552	80.9	106	18	US-09-461-580A-19	Sequence 19, Appli
7	550	80.6	823	20 <td>US-09-619-049-576</td> <td>Sequence 576, App</td>	US-09-619-049-576	Sequence 576, App
8	550	80.6	823	22 <td>US-60-167-217-10865</td> <td>Sequence 10865, A</td>	US-60-167-217-10865	Sequence 10865, A
9	550	80.6	823	22 <td>US-60-171-627-915</td> <td>Sequence 915, App</td>	US-60-171-627-915	Sequence 915, App
10	550	80.6	823	22 <td>US-60-173-464-8827</td> <td>Sequence 8827, Ap</td>	US-60-173-464-8827	Sequence 8827, Ap
11	550	80.6	823	22 <td>US-60-191-637-10901</td> <td>Sequence 10901, A</td>	US-60-191-637-10901	Sequence 10901, A
12	550	80.6	823	22 <td>US-60-191-681-8549</td> <td>Sequence 8549, Ap</td>	US-60-191-681-8549	Sequence 8549, Ap
13	517	75.8	286	16 <td>US-09-270-767-44940</td> <td>Sequence 44940, A</td>	US-09-270-767-44940	Sequence 44940, A
14	338.5	49.6	232	18 <td>US-09-461-580A-35</td> <td>Sequence 35, Appli</td>	US-09-461-580A-35	Sequence 35, Appli
15	336.5	49.3	183	18 <td>US-09-417-507-28899</td> <td>Sequence 28899, A</td>	US-09-417-507-28899	Sequence 28899, A
16	321	47.1	339	16 <td>US-09-248-796-19599</td> <td>Sequence 19599, A</td>	US-09-248-796-19599	Sequence 19599, A
17	321	47.1	339	22 <td>US-60-169-629-517</td> <td>Sequence 517, App</td>	US-60-169-629-517	Sequence 517, App
18	320.5	47.0	389	22 <td>US-60-167-217-10557</td> <td>Sequence 10557, A</td>	US-60-167-217-10557	Sequence 10557, A
19	320.5	47.0	389	22 <td>US-60-173-464-8533</td> <td>Sequence 8533, Ap</td>	US-60-173-464-8533	Sequence 8533, Ap
20	317.5	46.6	355	22 <td>US-60-167-217-10557</td> <td>Sequence 10557, A</td>	US-60-167-217-10557	Sequence 10557, A
21	317.5	46.6	355	22 <td>US-60-173-464-8533</td> <td>Sequence 8533, Ap</td>	US-60-173-464-8533	Sequence 8533, Ap
22	317	46.5	267	18 <td>US-09-461-580A-3</td> <td>Sequence 3, Appli</td>	US-09-461-580A-3	Sequence 3, Appli
23	309.5	45.4	107	18 <td>US-09-461-580A-20</td> <td>Sequence 20, Appli</td>	US-09-461-580A-20	Sequence 20, Appli
24	305.5	44.8	230	22 <td>PCT-US00-02237-9</td> <td>Sequence 9, Appli</td>	PCT-US00-02237-9	Sequence 9, Appli
25	305.5	44.8	230	22 <td>US-60-117-905-9</td> <td>Sequence 9, Appli</td>	US-60-117-905-9	Sequence 9, Appli
26	304.5	44.6	257	22 <td>US-60-095-827-17</td> <td>Sequence 17, Appli</td>	US-60-095-827-17	Sequence 17, Appli
27	304	44.6	128	18 <td>US-09-461-580A-10</td> <td>Sequence 10, Appli</td>	US-09-461-580A-10	Sequence 10, Appli
28	304	44.6	272	18 <td>US-09-461-580A-2</td> <td>Sequence 2, Appli</td>	US-09-461-580A-2	Sequence 2, Appli
29	304	44.6	336	18 <td>US-09-461-580A-11</td> <td>Sequence 11, Appli</td>	US-09-461-580A-11	Sequence 11, Appli
30	303.5	44.5	284	16 <td>US-09-248-796-19606</td> <td>Sequence 19606, A</td>	US-09-248-796-19606	Sequence 19606, A
31	303.5	44.5	284	16 <td>US-60-096-409-19606</td> <td>Sequence 20918, A</td>	US-60-096-409-19606	Sequence 20918, A
32	290	42.5	313	22 <td>US-09-248-796-20918</td> <td>Sequence 20918, A</td>	US-09-248-796-20918	Sequence 20918, A
33	290	42.5	313	22 <td>US-60-096-409-19606</td> <td>Sequence 57, Appli</td>	US-60-096-409-19606	Sequence 57, Appli
34	278.5	40.8	381	22 <td>US-60-117-905-57</td> <td>Sequence 3938, A</td>	US-60-117-905-57	Sequence 3938, A
35	267.5	39.2	358	18 <td>US-09-417-507-3938</td> <td>Sequence 18, Appli</td>	US-09-417-507-3938	Sequence 18, Appli
36	263.5	38.6	107	18 <td>US-09-461-580A-16</td> <td>Sequence 16, Appli</td>	US-09-461-580A-16	Sequence 16, Appli
37	245.5	36.0	223	1	PCT-US99-29950-188	Sequence 188, App
38	245.5	36.0	223	19 <td>US-09-591-316-189</td> <td>Sequence 189, App</td>	US-09-591-316-189	Sequence 189, App
39	242	35.5	106	18 <td>US-09-461-580A-15</td> <td>Sequence 15, Appli</td>	US-09-461-580A-15	Sequence 15, Appli
40	235	34.5	106	18 <td>US-09-461-580A-14</td> <td>Sequence 14, Appli</td>	US-09-461-580A-14	Sequence 14, Appli
41	200.5	29.4	488	16 <td>US-09-248-796-19601</td> <td>Sequence 19601, A</td>	US-09-248-796-19601	Sequence 19601, A
42	200.5	29.4	488	22 <td>US-60-096-409-19601</td> <td>Sequence 19601, A</td>	US-60-096-409-19601	Sequence 19601, A
43	192.5	28.2	131	18 <td>US-09-461-580A-17</td> <td>Sequence 17, Appli</td>	US-09-461-580A-17	Sequence 17, Appli
44	187	27.4	86	18 <td>US-09-461-580A-21</td> <td>Sequence 21, Appli</td>	US-09-461-580A-21	Sequence 21, Appli
45	184.5	27.1	156	18 <td>US-09-417-507-38675</td> <td>Sequence 38675, A</td>	US-09-417-507-38675	Sequence 38675, A

ALIGNMENTS

RESULT 1
US-09-461-580A-9
; Sequence 9, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461, 580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-9

Query Match 100.0%; Score 682; DB 18; Length 128;
Best Local Similarity 100.0%; Pred. No. 5.6e-76;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 IVLTGAGVSVCIGIPDFRSRDIYARLAVDPDLPDQAMEDIYFKDRPFFKKEI 60
|||||

Db 1 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 60
QY 61 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 120
Db 61 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 120
QY 121 KVDCEAVR 128
Db 121 KVDCEAVR 128

RESULT 2
US-09-461-580a-4
; Sequence 4, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580a-4

Query Match 100.0%; Score 682; DB 18; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.4e-75;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 60
Db 34 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 93
QY 61 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 120
Db 94 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 153
QY 121 KVDCEAVR 128
Db 154 KVDCEAVR 161

RESULT 3
US-09-461-580a-12
; Sequence 12, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580a-12

Query Match 100.0%; Score 682; DB 18; Length 327;

Best Local Similarity 100.0%; Pred. No. 2.1e-75;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 60
Db 34 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 93
QY 61 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 120
Db 94 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 153
QY 121 KVDCEAVR 128
Db 154 KVDCEAVR 161

RESULT 4
US-09-461-580a-1
; Sequence 1, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580a-1

Query Match 100.0%; Score 682; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 6.5e-75;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 60
Db 249 IVLGAGVSVSCGIPDFSRDGIYARLAVDPDDLPPQAMDIIEYFRKDPPEFFAKEI 308
QY 61 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 120
Db 309 YPGQFQPSLCHKFTALSDKEGKLLNNTYQNTIDTLEQVAGIORIIOCHGSFATASCLICKY 368
QY 121 KVDCEAVR 128
Db 369 KVDCEAVR 376

RESULT 5
US-09-461-580a-26
; Sequence 26, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 737
; TYPE: PRT

PRIOR APPLICATION NUMBER: 60/175,763

RESULT 9
US-60-171-627-915
; Sequence 915, Application US/60171627

```

: GENERAL INFORMATION:
: APPLICANT: Vandell, Mark
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
: TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
: TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF
: FILE REFERENCE: C1000177
: CURRENT APPLICATION NUMBER: US/60/171,627
: CURRENT FILING DATE: 1999-12-23
: NUMBER OF SEQ ID NOS: 2237
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 915
: LENGTH: 823
: TYPE: PRT
: ORGANISM: Drosophila
US-60-171-627-915
```

```

Query Match          80.6%; Score 550; DB 22; Length 823;
Best Local Similarity 75.8%; Pred. No. 1.5e-58;
Matches 97; Conservative 20; Mismatches 11; Indels 0; Gaps 0;
```

```

OY 1 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDPQAMFDIEFKRDPPEFFAKAI 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 225 IVLTGAGVSVSCGIPDFRSTNGIYARLAHDPDLPDPQAMFDINFKRDPPEFFAKAI 284
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 61 YPGFOPSLCHKFTALSDKEGKLLRNTYQNTIDLEOVAGIORIILQCHGSFATASCLICKY 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 285 YPGFOPSPCHRFITKMETKGLLRNTYQNTIDLEOVAGIORIVIECHGSFSTASCTKCRF 344
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 121 KVDCAVR 128
    | : : : |
DB 345 KCNADALR 352
```

```

RESULT 10
US-60-173-464-8827
: Sequence 8827, Application US/60173464
: GENERAL INFORMATION:
: APPLICANT: Li, Peter W.D.
: TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
: TITLE OF INVENTION: THEREOF
: FILE REFERENCE: C1000173
: CURRENT APPLICATION NUMBER: US/60/173,464
: CURRENT FILING DATE: 1999-12-29
: NUMBER OF SEQ ID NOS: 30269
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8827
: LENGTH: 823
: TYPE: PRT
: ORGANISM: Drosophila
US-60-173-464-8827
```

```

Query Match          80.6%; Score 550; DB 22; Length 823;
Best Local Similarity 75.8%; Pred. No. 1.5e-58;
Matches 97; Conservative 20; Mismatches 11; Indels 0; Gaps 0;
```

```

OY 1 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDPQAMFDIEFKRDPPEFFAKAI 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 225 IVLTGAGVSVSCGIPDFRSTNGIYARLAHDPDLPDPQAMFDINFKRDPPEFFAKAI 284
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 61 YPGFOPSLCHKFTALSDKEGKLLRNTYQNTIDLEOVAGIORIILQCHGSFATASCLICKY 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 285 YPGFOPSPCHRFITKMETKGLLRNTYQNTIDLEOVAGIORIVIECHGSFSTASCTKCRF 344
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 121 KVDCAVR 128
    | : : : |
DB 345 KCNADALR 352
```

```

RESULT 11
US-60-191-637-10901
```

```

: Sequence 10901, Application US/60191637
: GENERAL INFORMATION:
: APPLICANT: Venter, J. Craig
: TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
: TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: C1000392
: CURRENT APPLICATION NUMBER: US/60/191,637
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 42660
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10901
: LENGTH: 823
: TYPE: PRT
: ORGANISM: DROSOPHILA
US-60-191-637-10901
```

```

Query Match          80.6%; Score 550; DB 22; Length 823;
Best Local Similarity 75.8%; Pred. No. 1.5e-58;
Matches 97; Conservative 20; Mismatches 11; Indels 0; Gaps 0;
```

```

OY 1 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDPQAMFDIEFKRDPPEFFAKAI 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 225 IVLTGAGVSVSCGIPDFRSTNGIYARLAHDPDLPDPQAMFDINFKRDPPEFFAKAI 284
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 61 YPGFOPSLCHKFTALSDKEGKLLRNTYQNTIDLEOVAGIORIILQCHGSFATASCLICKY 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 285 YPGFOPSPCHRFITKMETKGLLRNTYQNTIDLEOVAGIORIVIECHGSFSTASCTKCRF 344
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 121 KVDCAVR 128
    | : : : |
DB 345 KCNADALR 352
```

```

RESULT 12
US-60-191-681-8549
: Sequence 8549, Application US/60191681
: GENERAL INFORMATION:
: APPLICANT: Li, Peter W.D.
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
: TITLE OF INVENTION: USES THEREOF.
: FILE REFERENCE: C1000390
: CURRENT APPLICATION NUMBER: US/60/191,681
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 30973
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8549
: LENGTH: 823
: TYPE: PRT
: ORGANISM: DROSOPHILA
US-60-191-681-8549
```

```

Query Match          80.6%; Score 550; DB 22; Length 823;
Best Local Similarity 75.8%; Pred. No. 1.5e-58;
Matches 97; Conservative 20; Mismatches 11; Indels 0; Gaps 0;
```

```

OY 1 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDPQAMFDIEFKRDPPEFFAKAI 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 225 IVLTGAGVSVSCGIPDFRSTNGIYARLAHDPDLPDPQAMFDINFKRDPPEFFAKAI 284
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 61 YPGFOPSLCHKFTALSDKEGKLLRNTYQNTIDLEOVAGIORIILQCHGSFATASCLICKY 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 285 YPGFOPSPCHRFITKMETKGLLRNTYQNTIDLEOVAGIORIVIECHGSFSTASCTKCRF 344
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
OY 121 KVDCAVR 128
    | : : : |
DB 345 KCNADALR 352
```

```

RESULT 13
```

```

US-09-270-767-44940
; Sequence 44940, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44940
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44940

```

```

Query Match          75.8%; Score 517; DB 16; Length 286;
Best Local Similarity 74.4%; Pred. No. 4e-55;
Matches 90; Conservative 20; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 8 VSVSCGIPDRSRDGIYARLAVDPDLDPQAMFIEYFRKDRPPEFKAKETYPGQFQP 67
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1 VSVSCGIPDRSRDGIYARLAVDPDLDPQAMFIEYFRKDRPPEFKAKETYPGQFQP 60
QY 68 SLCHKATIALSDKRGKLLRNNTONIDPLEOVAGIQRIQCHGSFATASCLCKKVDCEAV 127
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 61 SPCHRRFIKMETKGLRLNNTONIDPLERVAGIQRIQCHGSFATASCLCKKRCNADAL 120
QY 128 R 128
DB 121 R 121

```

RESULT 14
US-09-461-580A-35

```

; Sequence 35, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 232
; TYPE: PRT
; ORGANISM: unknown
; FEATURE:
; OTHER INFORMATION: Deduced Amino Acid Sequence of cDNA
US-09-461-580A-35

```

```

Query Match          49.6%; Score 338.5; DB 18; Length 232;
Best Local Similarity 56.3%; Pred. No. 3.2e-33;
Matches 67; Conservative 21; Mismatches 26; Indels 5; Gaps 3;

```

```

QY 1 IYLTGAGVSVSCGIPDRSRDGIYARLAVDPDLDPQAMFIEYFRKDRPPEFKAKE 59
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88 ICLVAGISISAGIPDRSRDGIYANL--EKYHLPYPAIFELSYFKKHPEPFALAKE 145
QY 60 IYPGQFQPSLCHKFIALSDKRGKLLRNNTONIDPLEOVAGI--QRIQCHGSFATASCL 116
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 146 IYPGQFQFPICHKFIALSDKRGKLLRNNTONIDPLERVAGIQRIQCHGSFATASCL 204

```

```

RESULT 15
US-09-417-507-28899
; Sequence 28899, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PAT99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 28899
; LENGTH: 183
; TYPE: PRT
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (177)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are un
US-09-417-507-28899

```

```

Query Match          49.3%; Score 336.5; DB 18; Length 183;
Best Local Similarity 51.7%; Pred. No. 4e-33;
Matches 61; Conservative 30; Mismatches 22; Indels 5; Gaps 3;

```

```

QY 5 GAGVSVSCGIPDRSRDGIYARLAVDPDLDPQAMFIEYFRKDRPPEFKAKETYPG 63
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 4 GAGISISAGIPDRSRDGTGLYSMLA--FLDLPEDVDYFISFRENRPFYALARELAPG 61
QY 64 QFQPSLCHKFIALSDKRGKLLRNNTONIDPLEOVAGI--QRIQCHGSFATASCLCK 119
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 62 RYRPTTAHSFVKLLHDKGLLKHPTONIDPLERVAGIQRIQCHGSFATASCLCK 119

```

Search completed: February 16, 2001, 10:49:15
Job time: 456 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:43:17 ; Search time 76.21 Seconds
(without alignments)
146.718 Million cell updates/sec

Title: US-09-461-580a-12

Perfect score: 1732

Sequence: 1 VINILSEPPKRRKKRDINTI.....PLHISEDSSSPERTVPQDSS 327

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 segs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	479	27.7	257	21	Y73360	HTRM clone 2495790
2	354.5	20.5	212	20	Y48540	Human Breast tumor
3	138	8.0	205	19	W98718	H. pylori GHPO 808
4	104	6.0	696	21	Y13359	Amino acid sequenc
5	104	6.0	696	21	Y70671	Human PRO266 prote
6	97	5.6	689	17	Y49270	Lactoferrin sequen
7	97	5.6	708	17	W09343	Bovine lactoferrin
8	97	5.6	708	19	W57318	Bovine lactoferrin
9	97	5.6	708	19	W86022	Bovine lactoferrin
10	93	5.4	142	20	W98856	H. pylori GHPO 164
11	86.5	5.0	496	20	W96815	Smad7 protein used
12	86	5.0	268	18	W20357	H. pylori transmem

13	86	5.0	272	18	W21028	H. pylori transmem
14	84	4.8	341	20	Y35341	Chlamydia pneumonia
15	84	4.8	573	16	R77859	S. clavuligerus OR
16	82.5	4.8	495	20	W89551	Mouse Smad6 protei
17	82.5	4.8	572	20	Y29616	Human PI3K-alpha
18	82.5	4.8	621	20	Y29617	Human PI3K-beta p
19	82	4.7	276	21	Y49442	Human plasma Zn-al
20	82	4.7	870	20	Y35233	C. pneumoniae prot
21	82	4.7	2395	13	R29839	Deduced from Le1ys
22	81	4.7	267	19	W60071	Human endometrium
23	81	4.7	740	20	W61533	Human Fas-binding
24	80	4.6	945	20	Y35612	C. pneumoniae prot
25	80	4.6	1584	20	W99300	Human BAII protein
26	79.5	4.6	465	20	Y37026	Chlamydia trachoma
27	79.5	4.6	465	19	W61003	Streptococcus pneu
28	78.5	4.5	774	18	W29323	DNA polymerase wlt
29	78.5	4.5	1829	18	W29322	DNA polymerase wlt
30	78.5	4.5	2366	17	R95011	C. difficile toxin
31	78	4.5	392	20	W67891	Human secreted pro
32	78	4.5	392	21	Y73357	HTRM clone 2291241
33	78	4.5	2414	16	R84882	Transcription fact
34	78	4.5	2414	19	W40057	Cellular transcrip
35	77.5	4.5	2366	19	W68388	Clostridium diffic
36	77	4.4	399	18	W14162	R. rhodochrous pro
37	77	4.4	433	17	R88452	Retinoic acid rece
38	77	4.4	479	19	W04723	Aromatic acyl tran
39	77	4.4	2555	17	W81512	Drosophila fat fac
40	76.5	4.4	708	16	R71037	Draculin. Desmodu
41	76.5	4.4	1702	14	R38095	T. littoralis DNA p
42	76.5	4.4	2547	19	W81511	Drosophila fat fac
43	76	4.4	516	20	Y06396	Bacillus subtilis
44	76	4.4	516	20	Y07735	Human breast-speci
45	76	4.4	648	21	Y70018	Human Protease and

ALIGNMENTS

RESULT 1	Y73360	standard; Protein; 257 AA.
ID	Y73360	
AC	Y73360;	
XX		
DT	24-FEB-2000	(first entry)
XX		
DE	HTRM clone 2495790	protein sequence.
XX		
KW	HTRM: human transcriptional regulatory molecule; arteriosclerosis; AIDS;	
KW	arteriosclerosis; cirrhosis; cancer; leukemia; diabetes mellitus;	
KW	Addison's disease; multiple sclerosis; Rheumatoid arthritis; infection;	
KW	trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.	
XX		
OS	Homo sapiens.	
XX		
PN	W09957144-A2.	
XX		
PD	11-NOV-1999.	
XX		
PE	04-MAY-1999;	99WO-US09935.
XX		
PR	05-MAY-1998;	98US-0084254.
PR	07-AUG-1998;	98US-0095827.
PR	02-OCT-1998;	98US-0102745.
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;	
PI	Gerstlin EH, Patterson C, Baughn MR, Azimzai Y, Lu DM;	
XX		
DR	WPI: 2000-052941/04.	
DR	N-PSDB: 252445.	
XX		

PT New peptides useful for diagnosis, prevention and treatment of cancers
PT and immune disorders -

Claim .1; Page 120-121; 193pp; English.

CC
CC Y733-25-13369 are human transcriptional regulator molecule (HTRM) protein sequences. The HTRM protein and nucleotide sequences are useful for CC preventing or treating disorders associated with decreased expression or activity of HTRM which include cell proliferative disorders such as arteriosclerosis and cirrhosis; cancers including adenocarcinoma and leukemia; immune disorders such as AIDS, Addison's disease, diabetes mellitus; rheumatoid arthritis, multiple sclerosis, systemic lupus erythematosus, and myasthenia gravis; infections and trauma. Antagonists of the HTRM polypeptides are useful for treating or preventing disorders associated with increased expression or activity of HTRMs. HTRM polypeptides, their immunogenic fragments or oligopeptides are useful for screening libraries of compounds in drug screening techniques.

CC Polynucleotides encoding HTRM are useful for blocking the transcription of mRNA and regulating gene function by modulating the activity of HTRM. Vectors expressing HTRM or agonists can also be used to prevent or treat disorder associated with decreased HTRM expression. Antibodies which specifically bind HTRM and polynucleotides encoding HTRM are useful for diagnosing disorders associated with the expression of HTRM, particularly in assays that detect the expression of HTRM. Nucleotide sequences encoding HTRM may be useful to generate hybridization probes useful in mapping the naturally occurring genomic sequence and to detect differences in gene sequences among normal, carrier and affected individuals. Using diagnostic assays, cancer can be detected prior to the appearance of clinical symptoms and thereby progression of cancer can be prevented by aggressive treatment or preventive measures.

CC

SQ Sequence 257 AA;

Query Match	27.78;	Score 479;	DB 21;	Length 257;
-------------	--------	------------	--------	-------------

Matches 103; Conservative 50; Mismatches 73; Indels 20; Gaps 7;

```

OY      LTGAGVSVSCSIDPFRS--RDSIVARLAVDPDLPDPAMPDIEFFRRDDPPPEFAKEIY 94
Db      1 mvgagstpsipdfrtppsgslysl--qxdldypsaetelpfflmppfftlakely 58

OY      95 PGGFQPSLCHEKFIALSDKEGKLERNYNQNIDTLEQVAGT--QRILQCHGSFATASCLTK 152
Db      59 pgnkykpvuthyfilllndkgllillyeqnldglerwsglpaskiveahgfafasactvcvq 118

OY      153 YKVDCENAVRGCIENQVVPBCRPCADBPDLAIMRKEIFYEFGGNLPEOP--BRAKKYDKDEV 210
Db      119 rpfgedatradvmadrvpcvpc-----tygvkpdiafyfegbpjqrfliln---vdfpma 170

OY      211 DLLIVIGSSLKRPVALIPSSIPIHEVBQOLINREPLRHAF----DVELLGGODVIINE 265
Db      171 dlilllgtsleevfaalteavrsyprlllnrdlvgrplawhprsrdaqldgdvbnghves 230

OY      266 LCHRIG- 271
Db      231 lvelly 236

```

RESULT	2
Y48540	
ID	Y48540 standard; Protein; 212 AA

AC Y48540;

DT 08-DEC-1999 (first entry)

Human breast tumour-associated protein 1

KM Expressed sequence tag; EST; human; breast; cancer; gene therapy/
KM treatment; tumour; cytostatic; medicament.

OS Homo sapiens.

XX	DE19813839-A1
PN	
XX	

PD 23-SEP-1999.

PF 20-MAR-1998; 98DE-1013839.
YY

PR 20-MAR-1998; 98DE-1013839.
YY

PA (META-) METAGEN GES GENOMFORSCHUNG MBH
XY

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A, XY

DR WPI; 1999-528981/45
DR N-PSDB: 733613

XX	Human nucleic acid
PT	

PT tissue, useful for breast cancer therapy -
XX

PS Claim 25; 143; 188pp; German.
XX

CC This invention describes novel human nucleic acid sequences from tumor
CC breast tissue which have cytosynthetic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer. 148540-148617 represent protein fragments
CC encoded by the expressed sequence tags described in the method of the
CC invention.

Sequence 212 AA;

Query Match	20.58;	Score 354.5;	DB 20;	Length 212;
-------------	--------	--------------	--------	-------------

Matches 70; Conservative 34; Mismatches 39; Indels 7; Gaps 4;

[illegible]

RESULT	3
W98718	
ID	W98718 standard; Protein; 205 AA
vv	

AC W98718;

DT 31-MAR-1999 (first entry)
YY

DE H. pylori GHPO 808 protein.
yy

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis; non-ulcer dyspepsia

Helicobacter pylori

XX
PN W0984 3478-21

XX
PD 08-OCT-1998

XX
PF 01-APR-1998:

AA 29-JUL-1997; 97US-0902615
PR

CC or survival of nerve cells including Parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata. PRO265 can be used as an
 CC anti-thrombotic agent. PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair. PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.

XX Sequence 696 AA:

Query Match 6.0%; Score 104; DB 20; Length 696;
 Best Local Similarity 22.9%; Pred. No. 0.028;
 Matches 66; Conservative 32; Mismatches 94; Indels 96; Gaps 17;

OY 75 DIEYRKDPREFKAKYPCQFC-----PSLCHEFALS--DKEG 114
 DB 131 dleyiqad---fnllrdipgafgdnklevllndnllstlpnfvfyvplthldlr 186
 OY 115 KLRVYTONIDPLEQVAGIQRIL-----QCHGSFATASCLICKY-----KVD 156
 DB 187 nrklclpye-evleqpgyaeilllednpwdc-----tcqlslkewlenipknaigrvv 240
 OY 157 CEA---VRGDIENQVVR--CP---RCPADEPLAIMKEIVEFGENLEPOFRAMKYKD 208
 DB 241 ceaprlgskdlnettedqlcpklnrvdsslp-appagettlapplpfpftngqehd- 298
 OY 209 EVDLLIVGSS-----LKVRYPALPSSPIHEVPOILINREPLPHLFDVELL 256
 DB 299 ----atpysapnggclkipnwgiklrptaalatgssrnkp--lanslpcp----- 342
 OY 257 GDCDVIINELCHRLGGEYAKLCNPNVKSIEKPPRQKELVHSEL 304
 DB 343 ggc-----cdhlpsgslkmmcnrnvssladlkpk-----lsnvqel 380

RESULT 5
 ID Y70671 standard; Protein: 696 AA.

XX AC Y70671;
 XX DT 18-JUL-2000 (first entry)

XX DE Human PRO266 protein.

XX KW PRO266; UNQ233; dermatological; immunosuppressive; antiinflammatory;
 KW immunostimulant; antiaesthetic; antirheumatic; antiarthritic; virocidic;
 KW antiallergic; haemostatic; hepatotropic; antidiabetic; antinaemic;
 KW nephrotropic; neuroprotective; anticoagulant; immunological disorder;
 KW lung; pneumonia; skin; psoriasis; kidney; glomerulonephritis; arthritis;
 KW spondyloarthropathy; SLE; systemic lupus erythematosus; scleroderma;
 KW idiopathic inflammatory myopathy; anaemia; thrombocytopenia; diabetes;
 KW thyroiditis; Grave's disease; demyelinating disease; multiple sclerosis;
 KW Crohn's disease; hepatobiliary disease; hepatitis; asthma; human;
 KW graft-versus-host-disease.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers

FT FT Modified-site 17..23 /note= "N-myristoylation site"

FT FT Modified-site 18..22 /note= "N-glycosylation site"

FT FT Modified-site 30..34 /note= "Casein Kinase II phosphorylation site"

FT FT Modified-site 67..73 /note= "N-myristoylation site"

FT FT Modified-site 100..106 /note= "N-myristoylation site"

FT FT Modified-site 122..126 /note= "N-myristoylation site"

FT FT /note= "cAMP and cGMP-dependent protein kinase
 FT FT phosphorylation site"
 FT FT Modified-site 180..184 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 222..226 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 253..257 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 256..260 /note= "N-glycosylation site"
 FT FT Modified-site 302..308 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 328..334 /note= "N-myristoylation site"
 FT FT Modified-site 337..348 /note= "N-myristoylation site"
 FT FT Binding-site 337..348 /note= "Prokaryotic membrane lipoprotein lipid
 FT FT attachment site"
 FT FT Modified-site 343..349 /note= "N-myristoylation site"
 FT FT Modified-site 354..360 /note= "N-myristoylation site"
 FT FT Modified-site 363..367 /note= "N-glycosylation site"
 FT FT Modified-site 366..370 /note= "N-glycosylation site"
 FT FT Modified-site 416..420 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 465..471 /note= "N-glycosylation site"
 FT FT Modified-site 493..499 /note= "N-myristoylation site"
 FT FT Modified-site 573..577 /note= "N-myristoylation site"
 FT FT Modified-site 595..599 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 598..604 /note= "N-glycosylation site"
 FT FT Modified-site 603..609 /note= "N-myristoylation site"
 FT FT Modified-site 608..612 /note= "N-myristoylation site"
 FT FT Modified-site 646..650 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 646..650 /note= "cAMP and cGMP-dependent protein kinase
 FT FT phosphorylation site"
 FT FT Modified-site 655..659 /note= "N-glycosylation site"
 FT FT Modified-site 657..661 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 666..670 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site 693..697 /note= "Casein Kinase II phosphorylation site"
 FT FT Modified-site /note= "Casein Kinase II phosphorylation site"
 FT FT WO200015797-A2.
 FT FT 23-MAR-2000.
 FT FT 15-SEP-1999; 99WO-US21547.
 FT FT 17-SEP-1998; 98US-0100858.
 FT FT 17-SEP-1998; 98WO-US19437.
 FT FT (GENH) GENENTECH INC.
 FT FT Fong S, Goddard A, Gurney AL, Tumas D, Wood WT;
 FT FT WPI; 2000-271435/23.
 FT FT N-PSDB; 252205.
 FT FT Composition for treatment and diagnosis of immune related diseases e.g.
 FT FT Grave's disease comprises a PRO245, PRO217, PRO301, PRO266, PRO335,
 FT FT PRO331 or PRO326 polypeptide or its agonists or antagonists (preferably
 FT FT antibodies) -

XX Example 1; Fig 10; 201pp; English.

CC The present sequence is the human protein PRO266, encoded by UNQ233 cDNA,
CC designated as clone DNA37150. It is isolated from human foetal brain
CC tissue. PRO266 has significant homology to a SLIT protein, indicating
CC that it could be a leucine rich repeat protein.
CC It enhances or suppresses the infiltration of inflammatory cells into
CC tissues, proliferation of T-lymphocytes and modulates the immune
CC response. This sequence is useful for treatment of immune related
CC disorders, like SLE, rheumatoid/juvenile arthritis, spondyloarthritis,
CC systemic sclerosis (scleroderma), idiopathic inflammatory myopathies such
CC as dermatomyositis, Sjogren's syndrome, systemic vasculitis, sarcoidosis,
CC autoimmune haemolytic anaemia, thrombocytopenia, thyroiditis e.g. Grave's
CC disease, diabetes mellitus, immune-mediated renal disease e.g.
CC glomerulonephritis, demyelinating diseases such as multiple sclerosis and
CC Guillain-Barre syndrome, hepatobiliary diseases like hepatitis and
CC primary biliary cirrhosis, inflammatory and fibrotic lung diseases such
CC as inflammatory bowel disease (e.g. Crohn's disease), autoimmune or
CC immune-mediated skin diseases such as psoriasis, allergies like asthma,
CC immunological diseases of the lungs such as eosinophilic pneumonia and
CC transplantation associated diseases such as graft-versus-host-disease.

SQ Sequence 696 AA;

Query Match 6.0%; Score 104; DB 21; Length 696;
Best Local Similarity 22.9%; Pred. No. 0.028;
Matches 66; Conservative 32; Mismatches 94; Indels 96; Gaps 17;

QY 75 DIEFRKDPFFFAKEIYFGQFQ-----PSLCHKFIALS--DKEG 114
Db 131 diefgqd-----fallrdipgaqgdlnklevalldnllstpranvfyvpihldlrg 186
QY 115 KLRNTQNTDTEQVAGIORIL-----OCHGSFATASCLICKY-----KVD 156
Db 187 nrkltpye-evleqldpglelllednwdc-----lcdllskewlenpknaligrv 240
QY 157 CEA---VRGIDFNQVVR--CP---RCPADPEPLAIKPEIYFEGENLPEQFHRAMKDKD 208
Db 241 ceaprtlqgkldnettedqldcplknrvdsilp-appaqeetlaprpilpfpfktngqech- 298
QY 209 EVDLLIYIGSS-----LKVAPVALIPSSIEHVPQILINREPLPHLHFDVEL 256
Db 299 -----atpgsapnpgygtklipgnwqiklirptaiaatgsrnkp--lanslpcp----- 342
QY 257 GDCDVIINELCHRLGGEYARLCCNPVKLSETTEKPPRPORKEVHLSEL 304
Db 343 ggcg-----cdhpgsglkmncnnrvssladlkpk-----lsnvqel 380

RESULT 6
Y49270
ID Y49270 standard; protein; 689 AA.

AC Y49270;

DT 07-FEB-2000 (first entry)

DE Lactoferrin sequence.

KW Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
medicament; drink; feed; food poisoning.

XX Undentified.

XX JP11292789-A.

XX 26-OCT-1999.

XX 03-APR-1998; 98JP-0107167.

XX 03-APR-1998; 98JP-0107167.

XX (MEIP) MEIJI MILK PROD CO LTD.

DR WPI; 2000-018674/02.

PT Adhesion inhibitory compositions of enteropathogenic E. coli to cells -
used in medicines and foodstuffs

PS Disclosure; Fig 4; 7pp; Japanese.

CC The invention provides an adhesion inhibitory composition of
CC enteropathogenic Escherichia coli to cells. The composition contains a
CC lactoferrin degraded matter. The composition is useful as a medicament,
CC a drink and food or a feed. The composition is effective for prevention
CC and/or treatment of infectious food poisoning caused by enteropathogenic
CC E. coli.

SQ Sequence 689 AA;

Query Match 5.6%; Score 97; DB 21; Length 689;
Best Local Similarity 23.6%; Pred. No. 0.15;
Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDRPFFFAKEIYFGQFQPSLCHKFIALSCK-----EGK----- 115
Db 69 rdpkxlrpvaaelygtespqlhyavaavvkkgsnfdqldqgrkschtglrsagwllp 128
QY 116 --LKRNTQNTDTEQVAGIORILQCHGSFATASCLIC---KRYVCEAVRGDIFNQV 169
Db 129 mglirpyswteslepiqg-----avakfssacvpcidqyplncqlckgengenca 182
QY 170 PRCPADPEP-----LAIMKPEIYFEGENLPEQFHRAMKYDKDEVLD 212
Db 183 -----csrepyfgysgafkclqdgagdvafvketvf--enlpe-----kadrqyel 229
QY 213 LIVIGS-----SLKVRVALIPS 230
Db 230 lclnstrapydalfkechlaqvs 252

RESULT 7
W09343
ID W09343 standard; Protein; 708 AA.

AC W09343;

DT 18-MAR-1997 (first entry)

DE Bovine lactoferrin.

KW Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
transferrin; bactericidal activity; prostate; expression system; primer;
PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
alpha-amylase; Aspergillus oryzae; nutrition; bovine.

XX Bos taurus.

XX US5571691-A.

XX 05-NOV-1996.

XX 05-MAY-1989; 89US-0348270.

XX 28-OCT-1993; 93US-0145681.

XX 05-MAY-1989; 89US-0348270.

XX 28-SEP-1989; 89US-0413880.

XX 24-APR-1992; 92US-0873304.

XX 27-OCT-1992; 92US-0967947.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Conneely OM, Heaton DR, May GS, O'Malley BW;

XX WPI: 1996-505400/50.
 DR N-PSDB: T48031.
 XX
 PT New DNA encoding human lactoferrin, its natural alleles and
 PT substitution analogues - useful e.g. for preventing iron deficiency
 PT and as antiviral/antimicrobial agent
 XX
 PS Disclosure: Column 31-34; 92pp; English.
 XX
 CC This is the amino acid sequence of bovine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.
 CC Fragments of the protein are also known to have biological activity e.g.
 CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native
 CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
 CC peptide. The novel construct was then expressed in *Aspergillus oryzae*.
 CC The protein can be used for antibacterial and antiviral activities as
 CC well as an iron-carrying protein for nutritional or therapeutic
 CC applications.
 CC
 XX Sequence 708 AA:
 SQ
 Query Match 5.6%; Score 97; DB 17; Length 708;
 Best Local Similarity 23.6%; Pred. No. 0.16;
 Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;
 OY 81 KDRPFFFAKRIYGGQPSLCHKFIASDK-----EGK----- 115
 DB 88 rdpkylrpvaaeiylgtkspqthyyavavvkgsnfgldqgtrschtyglsrsgwllp 147
 OY 116 --LLRNTQNTDLEQVAGIQRILOCHGSEFATASCLIC---KYKVDCEAVRGDIFNOV 169
 DB 148 mglirpylswtesleplg-----avakffasvcpcldrqaypnlcqckgegenqca 201
 OY 170 PRCPRCPADP-----LAIKKPRIVFEGENLPQGFHRAMKYDKDEVDL 212
 DB 202 -----cssrepyfygsgafkclqdgagdvafvketvif--enlpe-----kadrqyel 248
 OY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 lclnstrapvdafkchlaqvps 271
 RESULT 8
 ID W57318 standard; Protein: 708 AA.
 AC W57318;
 XX
 DT 10-AUG-1998 (first entry)
 XX
 DE Bovine lactoferrin.
 XX
 KW Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;
 KW Fe binding site; bacteria; bactericidal; milk.
 XX
 OS Bos taurus.
 XX
 US5766939-A.
 PD 16-JUN-1998.
 XX
 PE 30-MAY-1995; 95US-0453703.
 XX
 PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 PR 30-MAY-1995; 95US-0453703.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.

XX
 PI Conneely OM, Heaton DR, May GS, O'Malley BW;
 XX
 DR WPI: 1998-361705/31.
 DR N-PSDB: V30771.
 XX
 PT Plasmids containing human lactoferrin DNA - for recombinant
 PT production of the enzyme, especially fragments having bactericidal
 PT activity
 XX
 PS Example 13; Fig 14; 92pp; English.
 XX
 CC The present sequence represents bovine lactoferrin. A plasmid has been
 CC developed which is suitable for the expression of a human lactoferrin,
 CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
 CC host cell. The plasmid comprises a DNA sequence encoding a naturally
 CC occurring human lactoferrin protein or an iron-binding lobe of
 CC lactoferrin and further comprises transcriptional and translational
 CC regulatory elements capable of regulating the expression of the
 CC lactoferrin-encoding DNA sequence in the transformed host cell. The
 CC plasmid is useful for producing recombinant human lactoferrin proteins
 CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
 CC in milk and other secretory fluids. It is involved in iron transfer
 CC and delivery in mammals. It has been implicated as a resistance
 CC factor in suckled new born infants against enteric infections; the
 CC iron binding activity thought to be responsible for bactericidal action
 CC as the bound iron deprives microorganisms of the metal, and interferes
 CC with their growth and reproduction. The protein can be used for
 CC nutritional and therapeutic applications. It has not been previously
 CC possible to efficiently produce lactoferrin recombinantly.
 CC
 XX Sequence 708 AA:
 SQ
 Query Match 5.6%; Score 97; DB 19; Length 708;
 Best Local Similarity 23.6%; Pred. No. 0.16;
 Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;
 OY 81 KDRPFFFAKRIYGGQPSLCHKFIASDK-----EGK----- 115
 DB 88 rdpkylrpvaaeiylgtkspqthyyavavvkgsnfgldqgtrschtyglsrsgwllp 147
 OY 116 --LLRNTQNTDLEQVAGIQRILOCHGSEFATASCLIC---KYKVDCEAVRGDIFNOV 169
 DB 148 mglirpylswtesleplg-----avakffasvcpcldrqaypnlcqckgegenqca 201
 OY 170 PRCPRCPADP-----LAIKKPRIVFEGENLPQGFHRAMKYDKDEVDL 212
 DB 202 -----cssrepyfygsgafkclqdgagdvafvketvif--enlpe-----kadrqyel 248
 OY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 lclnstrapvdafkchlaqvps 271
 RESULT 9
 ID W86022 standard; Protein: 708 AA.
 AC W86022;
 XX
 DT 22-FEB-1999 (first entry)
 XX
 DE Bovine lactoferrin (hLf) polypeptide.
 XX
 KW Lactoferrin; hLf; recombinant; iron-binding; bovine; porcine; eyedrop;
 KW Aspergillus; animal food; therapeutic additive; iron transport; human;
 KW virucidal; bactericidal; contact lens; skin care; eardrop; mouthwash;
 KW chewing gum; toothpaste; preservative; anti-infection; nutrition.
 XX
 OS Bos taurus.
 XX
 US5849881-A.

```

XX 15-DEC-1998.
PD
XX
XX 30-MAY-1995; 95US-0456106.
PF
XX
XX 28-OCT-1993; 93US-0145681.
PR
XX 05-MAY-1989; 89US-0348270.
PR
XX 24-APR-1992; 92US-0873304.
PR
XX 27-OCT-1992; 92US-0967947.
PR
XX 30-MAY-1995; 95US-0456106.
XX
XX (CONN/) CONNELLY O. M.
PA (HEAD/) HEADON D. R.
PA (MAYG/) MAY G. S.
PA (OMAL/) O'MALLEY B. W.
XX
XX Connelly OM, Headon DR, May GS, O'Malley BW;
PI
XX WPI: 1999-069808/06.
DR
XX N-PSDB: V80376.
DR
XX
XX Recombinant lactoferrin - and lactoferrin fragments
PT
XX
XX Example 13; Columns 33-38; 93pp; English.
XX
XX This represents a bovine lactoferrin polypeptide. The invention relates
CC to production of recombinant lactoferrin using cDNA sequences of human,
CC bovine and porcine lactoferrin. Regions of the cDNA such as the iron-
CC binding sites can be used to produce recombinant lactoferrin, especially
CC human lactoferrin (hLF) polypeptide product. Aspergillus host cell
CC transformed with an expression vector comprising a hLF DNA sequence, a
CC promoter, a translation initiation sequence and transcription and
CC translation termination sequences can be used for the production of the
CC recombinant hLF polypeptide. Recombinant lactoferrin polypeptides can be
CC used in human and animal foods, as therapeutic additives to enhance iron
CC transport and delivery, and for their virucidal and bactericidal
CC properties, as additive for eye drops, contact lens care solutions, skin
CC care products, ear drops, mouthwashes, chewing gum and toothpaste, as
CC preservatives for the above products, as therapeutic anti-infection
CC agents, as nutritional supplements and as sources of amino acids.
CC
XX
SQ Sequence 708 AA;

Query Match 5.6%; Score 97; DB 20; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.16;
Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

OY 81 KDRPFKFAKEIYPCGFOPSLCHKFFALSDK-----EGK----- 115
DB 88 rdykrlrpyaaelytkesqptlyavavvksnfgldqgkrkschqglgrsagwlp 147
OY 116 --LLRNYTONIDTLEQVAGIQRILQCHGSPATASCLIC---KYKYDCEAVRGDIFNQV 169
DB 148 mgllrpylswtealeplg-----avakffsascvpoldrgyplnclqckegengca 201
OY 170 PRCPRPADP-----LAIMEPIYFECENPPEQFHRAMKDKDEVDL 212
DB 202 -----csstrepfygysgafkclqdgagdvafvketvfv--enlpe-----kardqyel 248
OY 213 LIVIGS-----SLKVRPALIPS 230
DB 249 lclnmsrpydafkechlaqvps 271

RESULT 10
ID W98856 standard; Protein: 142 AA.
XX
AC W98856;
XX
DT 31-MAR-1999 (first entry)
XX

```

```

DE H. pylori GHPO 1640 protein.
XX
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX peptic ulcer disease.
XX
XX Helicobacter pylori.
XX
XX W09843478-A1.
XX
XX 08-OCT-1998.
XX
XX 01-APR-1998; 98WO-US06371.
XX
XX 29-JUL-1997; 97US-0902615.
XX 01-APR-1997; 97US-0833457.
XX 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleantious H, Miller C, Oomen RP, Tomb J;
XX
XX WPI: 1998-542293/46.
XX N-PSDB: X14575.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
XX Claim 8; Page 1928-1929; 2054pp; English.
XX
XX This sequence represents a Helicobacter pylori GHPO protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX
XX
SQ Sequence 142 AA;

Query Match 5.4%; Score 93; DB 19; Length 142;
Best Local Similarity 22.6%; Pred. No. 0.043;
Matches 35; Conservative 30; Mismatches 50; Indels 40; Gaps 8;

OY 124 IDTLEQVAGIQRILQCHGSPATASCLICRYKVDCEAVRGDIFNQVPRCPADPEPLAI 183
DB 5 teslek-asig-vlece-----ncsmtyldrd-----ynreceicpcyckkpvrl 48
OY 184 M-----KPEIVFGEN-----LPEQFHRAMKYDKDEVDLLVIGSSLKVRPALIPSS 231
DB 49 vatsyyqksevfyfvsnftgplfipfltlfkglvkvksewefaelannll-----1 98
OY 232 IPHEVPO--LLINREPLPHLPVDELIGDCDVIIN 264
DB 99 fhndiqekillnnkldhyrleidelekeltisyn 133

RESULT 11
ID W96815 standard; Protein: 496 AA.
XX
AC W96815;
XX
XX 21-APR-1999 (first entry)
XX
XX Smad7 protein used to raise antibodies.
XX
XX Smad7; Smad7; Smad7 minimal promoter; TGF-beta activity;
XX cell growth modulation; cancer treatment; fibrosis; Smad6;
XX lung cancer; eye defect reduction; mammalian embryo.
XX

```

[illegible]

KW		binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM		dual ulcer disease; chronic gastritis; diagnosis; envelope.
XZ		
XX	OS	Helicobacter pylori.
PX	PN	MO6640893-AI.
PD	PP	19-DEC-1996.
PF	PE	06-JUN-1996; 96WO-U509122.
PR	PR	01-APR-1996; 96US-0630405.
PA	PR	07-JUN-1995; 95US-0487032.
PI	PA	(ASTR) ASTRA AB.
DR	PI	Berglindh OT, Smith D, Mellgaerd BL;
DN	DR	WPt: 1997-052306/05.
DT	DN	N-PADB: T67549.
PS	PT	Helicobacter pylori nucleic acid sequences and related polypeptide(s) - useful for vaccines to treat or prevent H. pylori infection, and to detect Helicobacter
XX	PS	Claim 73; Page 548; 1481pp; English.
CC	XX	This sequence is a H. pylori transmembrane protein likely to contain at least two membrane spanning regions.
CC	CC	The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds,
CC	CC	useful as potential H. pylori life cycle activators or inhibitors.
CC	CC	The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC	CC	and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified CC and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.
XX	CC	
SQ	XX	Sequence 268 AA;
		Query Match 5.0%; Score 86; DB 18; Length 268; Best Local Similarity 23.6%; Pred. No. 0.58; Matches 34; Conservative 21; Mismatches 49; Indels 40; Gaps
QY		120 YTONIDTLEQVAGIQRILOCHGSFATASCLICRY-----KVD-----C 157 : : : : : : : : : : : : : : : DQ yegqik--ekavnsq-----gkftpsylfgayegiknkidtnyvkpncksvneml 82
QY		158 EAIVGDI--ENQVPKPCRPADPPLAIIMKEIFYESEN.PEQFRAMKDCKDEVLLTY 215 : : : : : : : : : : : : : : : DQ ekltgdiddfnalygklipnjkdnfklkpqspfinynta-----nekeknrvl 133
QY		216 IGSSIAKVPAVALIPSSIPHEVPOI 239 : : : : : : : : : : : : : : : DQ lsrissqketcpplisiksnyshl 157
ID		RESULT 13
XX		W21028
AC		W21028 standard; Protein; 272 AA.
XX		W21028;
DT		23-JUL-1997 (first entry)
XE		H. pylori transmembrane protein, hp6pi0723orf13.
XX		

[illegible]

DT 13-NOV-1995 (first entry)
 XX
 DE S. clavuligerus ORF2 product.
 XX
 KW Clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;
 XX acetoxyhydroxyacid synthase.
 XX Streptomyces clavuligerus.
 XX
 PN CA2108113-A.
 XX
 PD 09-APR-1995.
 XX
 PF 08-OCT-1993; 93CA-2108113.
 XX
 PR 08-OCT-1993; 93CA-2108113.
 XX
 PA (UTAL-) UNIV ALBERTA.
 XX
 PI Aldoo KA, Jensen SE, Paradkar AS;
 XX
 DR WPI: 1995-207301/28.
 DR N-PSDB; Q91580.
 XX
 PS Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
 PT biosynthesis of the antibiotic in Streptomyces hosts which do not
 PT naturally produce clavulinate
 XX
 PS Claim 24; Fig.11; 41pp; English.
 XX
 CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),
 CC extending downstream from pcbC, included 10 ORFs encoding the
 CC enzymes required for clavulinate biosynthesis. The ORF2 product
 CC (R77859) showed a high degree of similarity to acetohydroxyacid
 CC synthases from various sources.
 XX
 SQ Sequence 573 AA;

Query Match 4.8%; Score 84; DB 16; Length 573;
 Best Local Similarity 21.0%; Pred. No. 2.8;
 Matches 43; Conservative 24; Mismatches 68; Indels 70; Gaps 8;
 QY 8 PPKRRKKRDINTIED-----AVKLLQECKRIIVLTGAGVSCGIPDERS----- 52
 DB 186 ppantpakpygvvadgwgkaadgaallaeakhpvlvgaalrsgavpatrelaqrlnl 245
 QY 53 -----RDGIYARLAVDFPDLPPQAMDEYFRKDPPE 86
 DB 246 pvltyiakgvlpvghelnvgavtgmqdl-----lnfpal---qumf-----apv 288
 QY 87 -----FKEFAKEIYPGOPPSLCHKFTALSDKEGKLBNYQNTIDTLEQVAGIQRILOC 139
 DB 289 dlvlvgygyaeadlrpsmqkglekktvrlsptvnpiprvpavdvclvalave----- 344
 QY 140 HGSFATASCLICKYKVDCEAVRGDI 164
 DB 345 hferatas-fgakqrhndleplrari 368

Search completed: February 16, 2001, 10:43:20
 Job time: 102 sec

regulatory protein SIR2 - yeast (*Kluyveromyces marxianus* var. *lactis*)

N:Alternate names: silent information regulator 2

C:Species: *Kluyveromyces marxianus* var. *lactis*, *Candida sphaerica*

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999

C:Accession: A56048; S36616

R:Chen, X.J.; Clark-Walker, G.D.

Mol. Cell. Biol. 14, 4501-4508, 1994

A:Title: sir2 mutants of *Kluyveromyces lactis* are hypersensitive to DNA-targeting drugs.

A:Reference number: A56048; MUID:94277055

A:Accession: A56048

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-670 <CH2>

A:Cross-references: EMBL:X74569; NID:9397349; PIDN:CAA52661.1; PID:9397350

C:Superfamily: regulatory protein SIR2

C:Keywords: DNA binding; nucleus; transcription regulation

Query Match 35.0%; Score 606.5; DB 2; Length 670;

Best Local Similarity 39.4%; Pred. No. 1.7e-43;

Matches 148; Conservative 56; Mismatches 107; Indels 65; Gaps 12;

11 RKRKRDITIEAVKLLDCKKIIVLTGAGVSCGIDPFRSDGIVARLAVDFPDLDP 70

121 RIRLNFHTIDFVAKLTKAKKIIVLTGAGISTSLGIDPFRSEGFYSKLG-DL-GLNDVP 348

71 QAMFIEFRKDRPFREFAKEIYGPQPSLCHKFIALSDKEGKILRNNTONIDLEOV 130

349 QDVFELEVFTEPDSVFNIAHNVLPENNYSPHLSFIKIDQKILRNNTONIDLESY 408

131 AGI--QRILQCHGFATASCLICIKYKVDCEAVRGDIFNOVPRCPRC-----PAD 178

409 AGVEPEKMQCHGSFATASCVTHMKIQGERIFPNIRNLQPLPCYCKSKRLFEFKTMD 468

179 EPLA-----IMKEPIVFEENLEPEQFHRAMKYKDEVDLLIVGSS 218

469 BELAGDEDDMDHGRSVPSFGLKPDITTFGEALPSKFRRLREVLDGDLICIGT 528

219 SLKVPVALIPSSIPHEVQILINREPLPHLHFDVLLGDCDVIINELCHRLGGE----- 273

529 SLKVPVSEIVMIRAHNPQVILINKDPVKAHFDLSLGLCDVAVLAOKGMPIDPHDN 588

274 YAKLCNPNV-----KLSEITEKPRPQ-KELVHL-----SELPTPLHISED- 314

589 WNKLT-KNVVFQSEEVERGYKVPPLINESPALEAEKHLPLQOSTAALT-PVSLSDS 646

315 -----SSSPERTVPOD 325

647 PERSSSSSPQPTPOD 662

RESULT 3

S59698

HST1 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein O1161; protein YOL068c

C:Species: *Saccharomyces cerevisiae*

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Jun-2000

C:Accession: S59698; S62108; S66761; S70376

R:Derbyshire, M.K.; Weinstein, K.G.; Strathern, J.N.

submitted to the EMBL Data Library, September 1995

A:Description: HST1, a new member of the SIR2 family of genes.

A:Reference number: S59698

A:Accession: S59698

A:Molecule type: DNA

A:Residues: 1-503 <DER>

A:Cross-references: EMBL:L47120; NID:9972892; PIDN:AAB38430.1; PID:9972893

R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D.

submitted to the EMBL Data Library, October 1995

A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in sil

A:Reference number: S62108

A:Accession: S62108

A:Molecule type: DNA

A:Residues: 1-503 <BRA>

A:Gene: SPDB:SPBC16D10.07c

A:Cross-references: EMBL:U39041; NID:91055019; PIDN:AAA81033.1; PID:91055020

R:Alexandraki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66756

A:Accession: S66761

A:Molecule type: DNA

A:Residues: 1-503 <ALE>

A:Cross-references: EMBL:Z74810; NID:91419890; PIDN:CAA9078.1; PID:91419891; MIPS:YO

A:Experimental source: strain S288C

R:Derbyshire, M.K.; Weinstein, K.G.; Strathern, J.N.

Yeast 12, 631-640, 1996

A:Title: HST1, a new member of the SIR2 family of genes.

A:Reference number: S70376; MUID:96405912

A:Accession: S70376

A:Molecule type: DNA

A:Residues: 1-503 <DEM>

A:Cross-references: EMBL:L47120; NID:9972892; PIDN:AAB38430.1; PID:9972893

C:Genetics:

A:Gene: SGD:HST1

A:Cross-references: SGD:S0005429; MIPS:YOL068c

A:Map position: 15L

C:Superfamily: regulatory protein SIR2

C:Keywords: transmembrane protein

F:1-17/Domain: transmembrane #status predicted <TM1>

F:402-418/Domain: transmembrane #status predicted <TM2>

Query Match 34.0%; Score 589; DB 2; Length 503;

Best Local Similarity 39.1%; Pred. No. 3.7e-42;

Matches 127; Conservative 59; Mismatches 99; Indels 40; Gaps 5;

11 RKRKRDITIEAVKLLDCKKIIVLTGAGVSCGIDPFRSDGIVARLAVDFPDLDP 70

161 RRLPNTFTIDFTLNNAKKIIVLTGAGVSTSLGIDPFRSEGFYSK--IRHGLDLP 238

71 QAMFIEFRKDRPFREFAKEIYGPQPSLCHKFIALSDKEGKILRNNTONIDLEOV 130

QDVFELEVFTEPDSVFNIAHNVLPENNYSPHLSFIKIDQKILRNNTONIDLESY 298

131 AGI--QRILQCHGFATASCLICIKYKVDCEAVRGDIFNOVPRCPRC----- 175

239 AGIDPKLVQCHGSFATASCVTHMKIQGERIFPNIRNLQPLPCYCKSKRLFEFKTMD 358

176 -----PADEPLAIKPEIIVFEENLEPEQFHRAMKYKDEVDLLIVGSSLAVR 223

359 NNTVGFNINENPILKSGVGLKPDITTFGEALPSFHTIRKIDLECDLLICIGTSLAVA 418

224 PYALIPSSIPHEVQILINREPLPHLHFDVLLGDCDVIINELCHRLGGEYA 275

419 PVSEIVNVPSSHVPQILINRDVTHAEFDNLNLCGCDVAVSLAKCHMDIPKKMODLK 478

276 KLCCNPVKLSEIT---EKPRPQKE 297

479 KIDVNCETIDKGTIKIKQPRKKQ 503

RESULT 4

T39571

probable regulatory protein sir2-like - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39571

R:Mood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.

submitted to the EMBL Data Library, March 1999

A:Reference number: 221864

A:Accession: T39571

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-471 <WOO>

A:Cross-references: EMBL:AL035637; PIDN:CAB38511.2; GSPDB:GN00067; SPDB:SPBC16D10.07c

A:Experimental source: strain 972h-; cosmid C16D10

C:Genetics:

A:Gene: SPDB:SPBC16D10.07c

A;Map position: 2
A;Introns: 53/1; 79/1; 106/3; 117/2; 263/1

Query Match	33.28;	Score 575.5;	DB 2;	Length 471;
Best Local Similarity	43.38;	Pred. No. 4.8e-41;		
Matches 135; Conservative	42;	Mismatches 94;	Indels 41;	Gaps 9

```

0Y 7 EPPRRKKKKDINTJEDAKLLLOEKKIIVLGAQVNSCGIPDRSRDGIATLAVDFPD 66
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 EVARRIKLPHNETEDVNNKKNNVVLGAISTISLGLDRSDSGFARIARH--G 168
0Y LPDQAMEDIYFPRKDRPFEPFKAEIYP--GQFQPSLCHKFIALSDKEGLLNTYQNI 124
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 LSEBSEMDDITFRFENPEITFTPARDLLPEINHTSPS--HAFIILLEKKNLSTLFTQNI 244
0Y 125 DTLEQVAGI--QRILOCHGSFATASCLICKKYKVDCEAVRGDIFNOQVYPRCPRC----- 175
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 245 DNLEKFTGJLSDNKLIOGHSFATATCIKCKKRVDSGSELYEDIRNQVRSYCECKGKPLKL 304
0Y 176 -----PADPDLA---IKKPEITVFGEENLPRQ-FHRAMKDYKDEVLD 213
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 305 RRVGONKKKEKHYFSDGDSESEDDLAQGIKMPDITFEGEALDPSPFNKVGSGCELEETDL 364
0Y 213 LIVGSSLKAVRPVALIIPSSIPHEVPQIILNEPLPHLFDVOLLGD--CDVILNELCRLG 271
    | | | | | | | | | : | | | | | | | | | | | | | | | | | | | |
Db 365 LICIGTSLKAVPVSSELISVIPPTPOIYISTRPVRHQFDVNFSPYCDMWIVIEICRAG 422
0Y 272 --GEYAKLCNRP 281
    | | | |
Db 425 WLNETLQALCDLP 436

```

RESULT 5
RGBS2
regulatory protein SIR2 - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein D7714; protein YDL042c; silent information regulator 2
C:Species: *Saccharomyces cerevisiae*
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C:Accession: S05891; S67575
R:Shore, D.; Squire, M.; Nasmyth, K.A.
EMBO J. 3, 2817-2823, 1984
A:Title: Characterization of two genes required for the position-effect control of yeast
A:Reference number: S05891; MUID:85126876
A:Accession: S05891
A:Molecule type: DNA
A:Residues: 1-562 <SHO>
A:Cross-references: EMBL:X01419; NID:94469; PIDD:CAA25667.1; PID:94470
R:Paulin, L.; Saren, A.M.; Laamanen, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67560
A:Accession: S67575
A:Molecule type: DNA
A:Residues: 1-562 <PAU>
A:Cross-references: EMBL:Z74090; NID:91431026; PIDD:CAA98600.1; PID:91431027; GSPDB:GN00
A:Experimental source: strain 5288C
C:Genetics:
A:Gene: SCD:SIR2; MIPS:YDL042c
A:Cross-references: SGD:S0002200; MIPS:YDL042c
A:Map position: 4L
C:Superfamily: regulatory protein SIR2
C:Keywords: DNA binding; transcription regulation

Query Match	32.8%;	Score 568;	DB 1;	Length 562;
Best Local Similarity	39.1%;	Pred. NO. 2.6e-40;		
Matches 127;	Conservative 48;	Mismatches 92;	Indels 58;	Gaps 7;

```

QY      1 VINIISSEPPKRRK-----KQIN-----TIEDAVKLLIQEKKIIVLTG 38
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db      205 VTNN--EQLPLAKQYVRLIKDQRAINKVCLTRLKLSNFTTIDHETQKHTARKIIVLTG 26
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY      39 AGVSVSCGIPDRSRDGIYARLAVDPDLDPQAMFDEYFKKDRDPFFKFAKEIYPGOF 98

```

[illegible]

```

RESULT      6
T18247
transcription regulatory protein - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18247
R:Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18831
A:Accession: T18247
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-331 <BAR>
A:Cross-references: EMBL:AL033503; NID:e1341066; PID:e1341074; PIDN:CAA2018.1
C:Genetics:
A:Note: Ca49C4.08c

```

Query Match	30.2%	Score 522.5	DB 2	Length 331
Best Local Similarity	36.7%	Pred. No. 9.5e-37		
Matches 117	Conservative 66	Mismatches 105	Indels 31	Gaps 8

Oy	19	TIIVAVKLLJOC---- <td>73</td>	73
Db	3	SLDILKVAARAVXNGKKVVEFNFGAISTGADIPFSPPTGLANLAK--LNLPEFAV	60
Oy	74	FDIEIFRDPPEPFKFAKEIYGOFOSLCKFTALSDKEGKLLRNTONIDTLEOAGI	133
Db	61	FDIDFFKEDPPEFTLAEELPGNFAPLTKHHFTKLLQDQSLKRVYTONIDTLERAGV	120
Oy	134	Q--RILOGSHFAFASCLCIKKYKDCAEVNRGDINQVAPRCPRCPADEPLAIKPEIVF	191
Db	121	EDKYIVEAHGSFASHHCVDCKHEMTTELTKLYMDKKIIPSGOHE-----GYKPRDLYVF	175
Oy	192	GENLPEQCFHRAKMDKODEVDLLIYSSGLKAVPALIPSSIPHEVQILINREPLPLHF	251
Db	176	GEGIPVKKFFEDLMEDDCEDVEVAIIVAGTSILVFEPSALPGEVYNNKCLRLVXNKEVGTFKH	235
Oy	252	-----DVLLEDDCVYIINELCHRLGGERYAKLCCNPVYALSETTEKPRPPQKELVHLSLP	306
Db	236	EPRKSDIITALHDCDIIVAEKLTCTLEGLP-----DKLNEYEEK---EKIKYSKAETKE	283
Oy	307	TPHHSIEDSSSPERTVPD	325
Db	284	TKMHEIEDKLEEAHLKED	302

RESULT 7
T40929
transcription regulator SIR2 homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C:Accession: T40929

Db 308 IYH-DLENSTLQSEHESADK 328

RESULT 10

H69827 conserved hypothetical protein yndz - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: H69827

R:Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallet
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koester, P.; Kolingstein, G.; Krog, S.; Kumano, M.; Kuita, K.; Lapius, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogilvara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schlach, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serio
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; M0ID:98044033

C:Accession: H69827

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-247 <KUN>

A:Cross-references: GB:299109; GB:AL009126; NID:g2633260; PIDN:CAB12804.1; PID:g2633300

A:Experimental source: strain 168

C:Genetics:

A:Gene: yndz

C:Superfamily: conserved hypothetical protein b1120

Query Match 20.2%; Score 349; DB 2; Length 247;

Best Local Similarity 32.7%; Pred. No. 3.3e-72;

Matches 87; Conservative 54; Mismatches 91; Indels 34; Gaps 9;

Db 1 IEDAVKLLQCKKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPPOAMFDIEYF 79

QY 80 RKDPRPFRAKEIY---PGQFQPSLCHKFTALSDKGLRNNTQNTIDTLEQVAGIQR 135

Db 54 LSYRLEWPFKEKELFQMKMGSEFNEGHLLAELEKQGVQIDFQNTIDGLKAKGSRH 113

QY 136 ILQCHSFATASCLCKKYKDCAVRGDITNOVYPRCPRADEPL--AIKKEIYFEGE 193

Db 114 VVELHESIQTAACPGACGARDLP---HLLEREVEPCTAAGNGDTCGTAKTDVVLFGD 169

QY 194 NLPEQFRAKRYK-DEVDLIYIGSSLKVPYALIP--SSIPHEVQIILNREPPLHL 250

Db 170 AV---MHFDLRYLKLQADLLIYIGSLEVAARFVPEDSLIPGMKVIINLEPT---- 222

QY 251 FDVELLDGDCVINECHRLGGEYAK 276

Db 223 -----YCDSLFDVYIHQIGEFAR 241

RESULT 11

A72370 regulatory protein, SIR2 family - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: A72370

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.
Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; M0ID:99287316

A:Accession: A72370

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-246 <ARN>

A:Cross-references: GB:AE001726; GB:AE000512; NID:g4980992; PIDN:AMD35575.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0490

C:Superfamily: conserved hypothetical protein b1120

Query Match 20.0%; Score 346; DB 2; Length 246;

Best Local Similarity 37.3%; Pred. No. 5.9e-72;

Matches 79; Conservative 41; Mismatches 76; Indels 16; Gaps 4;

Db 20 IEDAVKLLQCKKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPPOAMFDIEYF 79

QY 3 MKERLDLNSRLVLTGTAGISIPSGIPDFRGNIGYKYS-----QNVFDIDFF 53

Db 80 RKDPRPFRAKEIYFG-QFQPSLCHKFTALSDKGLRNNTQNTIDTLEQVAGIQR 137

QY 54 YSHPEEYFRFAKEGIFPMQAKPMLAHVLLAKLEKGLIEAVITQNTIDRLHORAGSKVI 113

Db 138 QCHGSFATASCLCKKYKDCAVRGDITNOVYPRCPRADEPLAIKKEIYFEGE 197

QY 114 ELHGQVEEYICVRCREKTYVEDYIKLESSDVPCLDCDN-----SLIRPNTVFEGNLPQ 168

Db 198 QFHRAMKYRDEVLDLIYIGSSLKVPYALIP 229

QY 169 DALREALGTSRASIMLVIGSSLVYVPAELP 200

RESULT 12

S54631

HSR3 protein - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein O2651; protein YOR025w

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999

C:Accession: S54631; S62109; S66891

R:de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54617

A:Accession: S54631

A:Molecule type: DNA

A:Residues: 1-447 <DEH>

A:Cross-references: EMBL:X67331; NID:g1041652; PIDN:CAA60741.1; PID:g829135

R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D

submitted to the EMBL Data Library, October 1995

A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in

A:Reference number: S62108

A:Accession: S62109

A:Molecule type: DNA

A:Residues: 1-447 <BRA>

A:Cross-references: EMBL:U39062; NID:g1055021; PIDN:AAA81034.1; PID:g1055022

A:Experimental source: strain IPH1 (S288C)

R:de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66877

A:Accession: S66891

A:Molecule type: DNA

A:Residues: 1-447 <DEW>

A:Cross-references: EMBL:X74933; NID:g1420134; PIDN:CAA9215.1; PID:e251955; PID:g142

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:HSR3

A:Cross-references: SGD:S0005551; MIPS:YOR025w

A:Map position: 15R

Query Match 19.3%; Score 333.5; DB 2; Length 447;

Best Local Similarity 29.6%; Pred. No. 1.4e-20;

Matches 107; Conservative 52; Mismatches 130; Indels 73; Gaps 15;

Tue Feb 27 15:47:00 2001

us-09-461-580a-12.rpr

Page 7

Query Match	16.58;	Score 285.5;	DB 2;	Length 247;
Best Local Similarity	29.68;	Pred. No. 7.9e-17;		
Matches 75; Conservative	48;	Mismatches 89;	Indels 41;	Gaps 9;

[illegible]

Search completed: February 16, 2001, 10:45:29
Job time: 230 sec

DB 291 RLRDINHHIDDEVAKEKIAKKLIVLIGAGISISLGLPDEKSSSEGFYSKLG-DL-GLNDP 348

[illegible]

DB 349 QDVESLEVFTEDESVFYNAHNVLPENMNSPLHSFIKMIQDKLLRNTYONIDNLESY 408
 QY 131 AGI--ORITQCHGSPATASCLICKYKVDCAVAGDIFNOVPRCPRC-----PAD 178
 DB 409 AGVEPEKMTQCHGSPATASCLICKYKVDCAVAGDIFNOVPRCPRC-----PAD 178
 QY 179 EPLA-----IKKPEIVFGENLPQGFHRAKMYKDEVDLIVIGS 218
 DB 469 EELADEDDDDMDHGRSVKSPGVLPDTEFGEALPSPKFRILREDVLCDDLICIGT 528
 QY 219 SLKVRVPLIPSSIPHEVPOILINREPLPHLHFDVLLGDCVYINELCHRGGE----- 273
 DB 529 SLKVRVSEIVNMIPAHVPOVILNKPVKAHFDLSLGLCDVALVAQKCGMDIPHDN 588
 QY 274 YKALCCNPV-----KLSEITEKPPRPQ-KELVHL-----SELPTPLHISED- 314
 DB 589 WKKL-KNKVFDSEVEERGVYKVPNLNESPALFELEBEKRLPLQOSTAALTP-PVSLSDS 646
 QY 315 -----SSPERIVPOD 325
 DB 647 PGRSSSSSPQPPPTQD 662

RESULT 2

HST1_YEAST STANDARD: PRT: 503 AA.
 AC P53685;

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HST1 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 1).
 GN HST1 OR YOL068C.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

RN Saccharomycetaceae; Saccharomyces.

[1]

SEQUENCE FROM N.A.

RC STRAIN-S288C / YPH1;

RA MEDLINE-96101589; PubMed-7498786;

RA Brachmann C.B., Sherman J.M., Devlane S.E., Cameron E.E., Pillus L.,

RA Boeke J.D.;

RT "The SIR2 gene family, conserved from bacteria to humans, functions

in silencing, cell cycle progression, and chromosome stability.";

RL Genes Dev. 9:2888-2902(1995).

[2]

SEQUENCE FROM N.A.

RC STRAIN-GRY 668;

RA MEDLINE-96405912; PubMed-8810037;

RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;

RT "HST1, a new member of the SIR2 family of genes.";

RL Yeast 12:631-640(1996).

[3]

SEQUENCE FROM N.A.

RA MEDLINE-97321807; PubMed-9178509;

RA Tzezema M., Katsoulou C., Alexandraki D.;

RT "Sequence analysis of a 3.2 kb segment from the left arm of yeast

chromosome XV reveals eight known genes and ten new open reading

frames including homologues of ABC transporters, inositol

phosphatases and human expressed sequence tags.";

RL Yeast 13:583-589(1997).

-1- FUNCTION: INVOLVED IN TELOMERIC SILENCING, AS OVEREXPRESSION

RESTORES SILENCING AT HMR IN SIR2 MUTANTS.

-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

-1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U39041; AAA81033.1; -
 DR EMBL: L47120; AAB38430.1; -
 DR EMBL: 274810; CAA99078.1; -
 DR SGD: S0005429; HST1
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 200 439
 FT ZN_FING 318 345
 FT SEQUENCE 503 AA; 57702 MW; 4CDE2799EA135ABB CRC64;

Query Match 34.0%; Score 589; DB 1; Length 503;
 Best Local Similarity 39.1%; Pred. No. 8.4e-42;
 Matches 127; Conservative 59; Mismatches 99; Indels 40; Gaps 5;

QY 11 RKKRNDINTEAVKVLQCKKTIIVTGAGVSCVCIPEPFRSDGIYALADFDLDP 70
 DB 181 RURLPENTIDHFTATLRNAKILVTAGVSTSLGIPFRSEGYSK--IRHLGLED 238
 QY 71 QAMFDIEVRKDPREFEKAETYPGQFQPSLCHKFIALSDKGLRNTYONIDLEQV 130
 DB 239 QDVFNIDIFLODPSPVFNNAHNVLPENMNSPLHSFIKMIQDKLLRNTYONIDNLESY 298
 QY 131 AGI--ORITQCHGSPATASCLICKYKVDCAVAGDIFNOVPRCPRC----- 175
 DB 299 AGIDPDKLVQCHGSPATASCLICKYKVDCAVAGDIFNOVPRCPRC----- 175
 QY 176 -----PADEPLAKMKPEIVFGENLPQGFHRAKMYKDEVDLIVIGSSIKVR 223
 DB 359 NNTVQTNINFNPSPLKSYGLVAPDMTFEGEALPSPKFRILREDVLCDDLICIGTSLKVA 418
 QY 224 PVALIPSSIPHEVPOILINREPLPHLHFDVLLGDCVYIN-----ELCHRGGEYA 275
 DB 419 PVSEIVNMVPSHPQILNRDMVTHAEFDNLNGLGCDVAVSLVAKKCHMDIPKKWQDLK 478
 QY 276 KLCNPVKLSEIT---EKPPPOKE 297
 DB 479 KIDVNCETEDKGTYYIKKQPRKKQ 503

RESULT 3

SIR2_YEAST STANDARD: PRT: 562 AA.
 AC P06700;

DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 GN SIR2 OR MARI OR YDL042C OR D2714.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

RN Saccharomycetaceae; Saccharomyces.

[1]

SEQUENCE FROM N.A.

RA MEDLINE-85126876; PubMed-6098447;

RA Shore D., Squire M., Nasmyth K.A.;

RT "Characterization of two genes required for the position-effect

control of yeast mating-type genes.";

RL EMBO J. 3:2817-2823(1984).

[2]

SEQUENCE FROM N.A.

RC STRAIN-S288C;

RA Paulin L., Saren A.M., Laamanen P.;

RT Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

-1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM

MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS

INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI

SILENCING. SUPPRESSES RDNA RECOMBINATION AND PROMOTES HISTONE

DEACETYLATION.

-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

-1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X01419; CAA25667.1; -
 DR EMBL: Z71781; CAA96447.1; -
 DR EMBL: Z74090; CAA98600.1; -
 DR PIR: S05891; KGBYS2.
 DR SDD: S0002200; SIR2.
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 254 498 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 372 399 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 562 AA; 63261 MW; 52E693753654586 CRC64;

Query Match 32.8%; Score 568; DB 1; Length 562;
 Best Local Similarity 39.1%; Pred. No. 5.4e-40;
 Matches 127; Conservative 48; Mismatches 92; Indels 58; Gaps 7;

QY 1 VINILSEPRKKR-----KGIN-----TIEDAVKLLQECKKIIVLTG 38
 DB 205 VTNV--EDPLAKQTVRLNDLQRAINKVLCRLSLNFTIDHFIQKIHAKIIVLTG 262
 QY 39 AGVSVSCGIDPFRSRDGIYARLAVDPDLPQAMFIEYFRKDPPEFKAKIYGOE 98
 DB 263 AGVSTSLGIDPFRSSEGFYSK--IKHGLDDPDQVFNIFMHPDSVFYNNANVLPPEK 320
 QY 99 QPSLCHEFLSDQEKLLNNTQNTIDLEQVAGI--QRILOCHGSFATASCLCKKVD 156
 DB 321 IYSLHSLFIMLQKGLKRLNRYTONIDNLESYAGISTDKLVQCHGSFATATCVTHMMLP 380
 QY 157 CEAVRGDIFNQVPRCPRC-----PADEP-----LAIMKP 186
 DB 381 GERFENKIRNLELPLCPYCKKREYPEGYNKVGVAASGSGSEPRPYILNSGYLKP 440
 QY 187 EIVFEGNLEPQFRAMKRYKDEVDLIVGSSLKVRPALIPSSIPHEVQIILNEPL 246
 DB 441 DITFEFGALPNKFKHSITREDILECDLICTGTSIKVAPSVIVMPSHVQVILNRPV 500
 QY 247 PHLHDEVLIGDQVITINELCHRIIG 271
 DB 501 KHAEDLSLGLYCDDIAMVAQAQKG 525

RESULT 4
 SIR2_CANAL STANDARD; PRT; 515 AA.
 AC 059923;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE REGULATORY PROTEIN SIR2.
 GN SIR2.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC anamorphic Saccharomycetales; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martín J., Johnson A.D.;
 RT "SIR2 gene from Candida albicans."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC DEACTIVATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF045774; AAC09304.1; -
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 234 453 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 352 379 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 515 AA; 57732 MW; 2E4DB259E9A9251F CRC64;

Query Match 32.3%; Score 559.5; DB 1; Length 515;
 Best Local Similarity 45.5%; Pred. No. 2.5e-39;
 Matches 116; Conservative 48; Mismatches 82; Indels 9; Gaps 3;

QY 20 IEDAVKLLQECKKIIVLTGAGVSVSCGIDPFRSRDGIYARLAVDPDLPQAMFIEY 79
 DB 224 LPDLISLRAKKIMVVTGAGISTSLGIDPFRSEKGLYNQLSK--LNLSDPQKVFDLQTF 281
 QY 80 RKDPRPEFKAKIYPOFQPSLCHKRIALSDKGLKRLNRYTONIDNLEQVAGI--QRI 137
 DB 282 MREGRLFTYTHAVLPLPDGKFSILHAFKLQDNKRLNRYTONIDNLEQVAGI--QRI 341
 QY 138 QCHGSFATASCLCKKYKDEAVRGDIFNQVPRCPRC-----PADEPLAIMEKPIYFG 192
 DB 342 QCHGSFPAKACVSCQGIAGKGIYNNHRRKQPCALCWKTKAPPHFGAKRTITFFG 401
 QY 193 ENLEPQFRAMKRYKDEVDLIVGSSLKVRPALIPSSIPHEVQIILNEPLPHLP 252
 DB 402 EDLPERFHTLMDKDLOQIDLEFLVIGTSLKVPASIIERYKPKILINRDPINRGN 461
 QY 253 VEILGDDVITINELC 267
 DB 462 LQLLGLCDDAVSYLC 476

RESULT 5
 SIR2_LEIMA STANDARD; PRT; 361 AA.
 AC Q25337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RP).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-96186914; PubMed-8635734;
 RA Yahiaoui B., Taibi A., Ouassil A.;
 RT "A Leishmania major protein with extensive homology to silent
 RT information regulator 2 of Saccharomyces cerevisiae."
 RL Gene 169:115-118(1996).
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; L40331; AAB06804.1; -
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
FT DOMAIN 148 151 POLY-ALA.
FT ZN-FING 152 179 C4-TYPE (POTENTIAL).
FT DOMAIN 271 278 POLY-SER.
FT DOMAIN 305 310 POLY-SER.
SQ SEQUENCE 381 AA; 41958 MW; 5A311630AA5D2365 CRC64;

Query Match 27.9%; Score 484; DB 1; Length 381;
Best Local Similarity 32.1%; Pred. No. 3.4e-33;
Matches 125; Conservative 55; Mismatches 101; Indels 108; Gaps 12;

OY 5 ISEPKRRKKRKINTIEDAVKLOE--CKKITVLTGAGVSVCGLPDRSRD-GIYARLA 61
DB 14 LGEP-----TVEGLARYIREKDVARRILVAGAGSVAGIPIDFRSSDTGIYAKLG 63
OY 62 VDFPLDPQAMFDEYFKDRPFKFAKE--IYGGOPSLCKHKFTALSDKEGKLLRN 119
DB 64 K--YNUDDPTDAFSLTLREKPEITYSTARLNLMPGHQPTAVHHFTRLQDEGRLLRC 121
OY 120 YTONIDTLEQVAGI--QRILQCHGSFATASCLICKYKDYCEAVRGDIFNQVVRGPRCPA 177
DB 122 CQNIDGLEKAGVSEPLLVEAHGSPAAACIECHTFSEIGNYLEAMSGTVSRCTGC- 180
OY 178 DEPLAIMPEIYFEGENLPEQFRAMKDYDKDEVLLIYGSLLAKRPVALIPSSIPHEVP 237
DB 181 ----GIYKRVNVFEGENLPDAFFDALHMDAPLAEVLIIIGTSMQVHPALLCPVVKPSIP 236
OY 238 QILINRE-----PLPHL----- 250
DB 237 RFLAMRERVGGLFRPPDPLDTIHDDAVAKGCRSSSSGSRPSASAREEGCTEDGSSS 296
OY 251 -----FDVELLGDGCVIINELCHRLG--GEYAKLC 278
DB 297 PNEVEDASTSSSDGYGQDYVAHPDYCRDVPFRGDCQENVLKLAEGLREALAKRC 356
OY 279 CNPVKLSITTEKPPROKELVHLSLPP 307
DB 357 ASPVR-HQLRHRRRMRRE---SELLPT 380

RESULT 6
HST2_YEAST STANDARD; PRT; 357 AA.
AC P53686;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST2 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 2).
GN HST2 OR YPL015C OR LPAC2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1;
RX MEDLINE=96101589; PubMed=7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Hall J., Ahmed A., Bussey H., Fortin N., Friesen J.D., Storms R.K.,
RA Vo D.H., Wang Y., Winnett E.;
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD FUNCTION IN TELOMERIC SILENCING, CELL CYCLE
CC PROGRESSION AND CHROMOSOME STABILITY.

```

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U39063; AAA81035.1; -
DR EMBL; U33335; AAB68090.1; -
DR SCD; S0005936; HST2.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
FT DOMAIN 22 250 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 143 173 C4-TYPE (POTENTIAL).
SQ SEQUENCE 357 AA; 39979 MW; ED281E5B8241A4D0 CRC64;

Query Match 24.9%; Score 431; DB 1; Length 357;
Best Local Similarity 34.4%; Pred. No. 8.3e-29;
Matches 111; Conservative 52; Mismatches 106; Indels 54; Gaps 10;

OY 32 KIIVLTGAGVSVCGLPFRS-RDGIVARLAVDFDLPQAMFDEYFKDRPFKFA 90
DB 26 KYIENVGAGISTSCIPDRSGTGLYNLA--RLKLYPEAVFVDVDFQSDPLPFTYLA 83
OY 91 KEIYGFOPSLCKHFIALSDREGKLLRNQTONIDTLEQVAGIQ--RILQCHGSFATASC 148
DB 84 KELYGNFRPSKFHYLLKFDQKDVLRVYQNIIDTLERQAGVKDDLIIEAHGSPAHCHC 143
OY 149 LICKKRVCEAVRGDIENQVVRGCRPADPEPLAIMKEIYFEGENLPEQFRAMKDYKD 208
DB 144 ICGKRVYPPQVFKSKLAEHPINDYVKC--DVCGEIYKPAIYFEGDLPDSFETWLNDS 201
OY 209 -----EVDLLIIGSSLSKVRVALIPSSIPHEVPQILINREPLPHLFDVEL 255
DB 202 WIREKITTSGRKRPQDPVIYVGTSLAYVPFASLPREIRKRVKRVLCN-----LET 251
OY 256 IGD-----CDVITNELCHRLGGEYAKLCNPVLSITTEKPPRPQ-----KE 297
DB 252 VQDFKANKRPFTLIY----HQYSDEFAEQIVELQMOEDFEKILTAQCGMDNSKEQLLE 307
OY 298 LVHLSLPPTPHISEDSSSPER 320
DB 308 IVH--DLENLSLDQSEHESADKK 328

RESULT 7
HST3_YEAST STANDARD; PRT; 447 AA.
AC P53687;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST3 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
GN HST3 OR TOR025W OR OR26.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1;
RX MEDLINE=96101589; PubMed=7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN-5288C / FY1679;
 RA de Haan M., Maarse A.C., Grievell L.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
 CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
 CC AND GENOMIC MAINTENANCE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; U39062; AAA81034.1; -;
 DR EMBL; X87331; CAA60741.1; -;
 DR EMBL; Z74933; CAA99215.1; -;
 DR SGD; S0005551; HST3.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT DOMAIN 50 314 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 195 223 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 447 AA; 50524 MW; 59CD5FBD712B7005 CRC64;

Query Match 19.3%; Score 333.5; DB 1; Length 447;
 Best Local Similarity 29.6%; Pred. No. 1.5e-20;
 Matches 107; Conservative 52; Mismatches 130; Indels 73; Gaps 15;

QY 16 DINTEDAVKVLQCKKIIVLTGAGVSYSCGIDPFRSDGIYARLAVD---FPDLPPDQA 72
 DB 38 DDEVLRVTVQLSRSRRLACTLGAGISCNAGIPFRSDGIYLVKKCCSQYWSIKSGRE 97
 QY 73 MEDIEYFRKDP--PFEFAKEIYPG--QFQPSLCHKFIALSDKGLRLRYTONIDPLE 128
 DB 98 MEDISLFRDPRKISIFAKFERKLSNVQAKPTHTKFIALDKRNLKRYTONIDPLE 157
 QY 129 QVAGIQ-----RILOCHGSFATASCLIC-----KYVDCEAVRGD 163
 DB 158 ESIGLTLNKRKLPLTSSSHMKNLVDVQHLGDLTLSTCKCFQTFPMRWYSRC-LARGE 216
 QY 164 IFNVOVRCPCPA-----DEPLAIKPELVEFGENLP--EQHRAKMD-- 206
 DB 217 L-----PLCPDCEALINKRLNEGKRTLSNVGILRPNIYLVGENHPSCEIITGCLNDII 271
 QY 207 KDEVDLLIVIGSSSLKVRVALIPSSIPHEVPQ-----ILINREPLP---HLHFDVLLG 257
 DB 272 KGNDFLLIKRTSLKVDGVQVKKKIHDRGLIITLVKTPIGSSWHGIIDYQIHS 331
 QY 258 DCDVIINELCHRLGGEYAKLCCNPVKLSEI-----TEKPRPQKELVHLSLPPTPLHI 311
 DB 332 DCDWVWFLESGI-PDFEKTQDQIKRLQLKREASDLKKQAKQD--SICPTPTPLRT 388
 QY 312 SE 313
 DB 389 AQ 390

RESULT 8
 HST4_YEAST STANDARD; PRT; 370 AA.
 ID HST4_YEAST
 AC P53688;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 4).
 GN HST4 OR YDR191W OR YD9346.03.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-5288C / AB972;
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP GENE NAME.
 RC STRAIN-GRY 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RA Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev 9:2888-2902(1995).
 CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST3 TO TELOMERIC SILENCING,
 CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
 CC AND GENOMIC MAINTENANCE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; Z48784; CAA88705.1; -;
 DR SGD; S0002599; HST4.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT DOMAIN 90 342 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 221 254 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 370 AA; 41765 MW; 86BB0238BFA914F1 CRC64;

Query Match 13.9%; Score 240.5; DB 1; Length 370;
 Best Local Similarity 27.2%; Pred. No. 6.7e-13;
 Matches 85; Conservative 53; Mismatches 113; Indels 61; Gaps 13;

QY 5 LSEPPKRRKKRDINTIEDAVKVLQCKKIIVLTGAGVSYSCGIDPFRSDGIYARL-AVD 63
 DB 70 LSKSORHMDADAGFISYA--LWYSKRMVYVSGAGISVAAGIDPFRSDEGIFSTVNGGS 126
 QY 64 FPDLPPDQAMDIE---YFRKDPPEPFAKEIYPGQFQPSLCHKFIALSDKGLRLN 119
 DB 127 GKDLFDYRVYGDSSMSLKFQNLVSLFRSLKNQPTFE---HEMNEFAFGRLRL 181
 QY 120 YTONIDILEQ-----VAGIORILOCHGSFATASCLIC-----KKVDCFA 159
 DB 182 YTONIDIGDTQLPHLSTVPLAKPIPTVQLHGSIKHMECKNKINIKFPDELKCD-- 238
 QY 160 VRGDFNVOVRCPCRADEPLAII-----KEIYFGENLPE--QFHAMKMD- 206
 DB 239 DKFDSREIFIPSCFQCEYETEVKRMAGIRSTGVGKLRPRVLIYNEVHDEGFIETIANN 298
 QY 207 -KDEVDLLIVIGSSSLKVRVALIPSSIPHEVPQ-----ILIN-----REPLPHLF--DVE 254
 DB 299 LKKRIDCLIIIVTSLKIPGVANICROFAKVAHANGIYLYNTSMPKRNVLDSLKFVDLV 358
 QY 255 ILGCDVIINEL 266
 DB 359 VLGDCCOHTSL 370

RESULT 9

FT CONFLICT 264 264 C -> Y (IN REF. 4).
 FT CONFLICT 273 273 A -> P (IN REF. 4).
 FT CONFLICT 281 281 G -> A (IN REF. 4).
 FT CONFLICT 291 291 S -> R (IN REF. 4).
 FT CONFLICT 297 297 S -> S (IN REF. 2).
 FT CONFLICT 340 340 G -> A (IN REF. 1).
 FT CONFLICT 418 418 I -> V (IN REF. 1).
 FT CONFLICT 439 439 H -> Y (IN REF. 1).
 FT CONFLICT 459 459 K -> R (IN REF. 1).
 FT CONFLICT 514 514 A -> R (IN REF. 1).
 SQ SEQUENCE 708 AA; 78056 MW; C6FD/FC15D68E93F CRC64;

Query Match 5.6%; Score 97; DB 1; Length 708;
 Best Local Similarity 23.6%; Pred. No. 1.4; Indels 72; Gaps 9;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDPAPFPKFAKEIYGFQFQPSLCHKFTALSDK-----EGK----- 115
 DB 88 RDPFKLRPAVAEIGTGESQTHYAVAVKKSINFQDLQGRKSGHTGLRSAGNIIIP 147
 QY 116 --LIRANTQNDILEQVAGIQRILQCHGSFATASCLIC---KYKDCAVRGDIENOV 169
 DB 148 MGLRPLSLWTESELEPLQG-----AVAKFESASCVCPCIDROAYPNLCQLCKGEGENQCA 201
 QY 170 PROCPADP-----LAIMKEIYFEGENLEPOGFRAMKVKDEVDL 212
 DB 202 -----CSREPEYFGSFAFCLODAGDVAFAKETTIV--ENLPE-----KADRDQYEL 248
 QY 213 LIVIGS-----SLKVRPALIPS 230
 DB 249 LCLNNSRAPVADAFKECHLAQVPS 271

RESULT 10
 ID Y021.METJA STANDARD; PRT; 375 AA.
 AC 060332;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL PROTEIN MJ0021.
 GN MJ0021.
 OS Methanococcus jannaschli.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschli.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: BELONGS TO THE ORGANIC RADICAL ACTIVATING ENZYMES
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U67461; AAB98003.1; -

DR TIGR; MJ0021; -
 DR INTERPRO: IPR001989; -
 DR PROSITE: PS01087; RADICAL_ACTIVATING; 1.
 KW Hypothetical protein; Iron.
 FT METAL 38 38 IRON (POTENTIAL).
 FT METAL 42 42 IRON (POTENTIAL).
 FT METAL 45 45 IRON (POTENTIAL).
 SQ SEQUENCE 375 AA; 43446 MW; 5AEA168D21383C8 CRC64;

Query Match 5.3%; Score 91; DB 1; Length 375;
 Best Local Similarity 19.3%; Pred. No. 2;
 Matches 73; Conservative 62; Mismatches 97; Indels 146; Gaps 21;

QY 5 LSEPPKRRKRD-----INTEDAVKLQCKKIIVLTGACVSCGIPDRSR--- 53
 DB 47 LSE--KRKNDDVIYANERLITVEEAIIEAKLC-----SSKGVITGCGNPLKINRTV 97
 QY 54 -----DGIYARLAVDPDLPDQAM-----FDIEYFRKDPPEFKAKETPGQF 98
 DB 98 KFLKALKKEDEFERHAHLYTT--PETVNEENLKLKADLDRIPLPTKIF--NEGIDEEY 153
 QY 99 -----QPSLCHKRI-----ALSDKEGLRLRYNTQNDILEQVAGIQRILQCH--- 140
 DB 154 IKFLCNKLNLCNKYIEDVGEIIPALPMNENILK-----LAAIDGIAKFMNINELEF 206
 QY 141 -----GSRFASCLICKKVDCEAVRGDIENOVPRCP 176
 DB 207 SEENYHELEKRGFMKDDVSNAGSEETALKVIEKFR-----GDLF---INVCPSVL 256
 QY 177 ADEPLAI-MKPEIYFEGENLEPOGFRAMKVKDEVDLLIVGSSLKVRPALIPSSIPHE 235
 DB 257 KD--AIQMRRLINRANNAKPYE-----VITEGGLLNGI----- 290
 QY 236 VPQILINREPLPHLFVVELDGCDV--INELCHRLGEGYAKLCNPVLSITEKRP 292
 DB 291 --MIFNEDDLKEM--AEILEENIEIEELIDK-----NICNPFLEIDIEEMK 335
 QY 293 R---POKELVHLSLEPPT 307
 DB 336 RQRPETFSAYISLEPPT 353

RESULT 11
 ID TRFL_BUBAR STANDARD; PRT; 708 AA.
 AC 077698;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE LACTOTRANSFERRIN PRECURSOR (LACTOFERRIN).
 GN LTF.
 OS Bubalus arnee bubalis (Domestic water buffalo).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bubalus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paramasivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
 RA Singh T.P.;
 RT "cDNA sequence of Buffalo lactoferrin.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RA Kathiravan S., Paramasivam M., Yadav S., Srinivasan A., Singh T.P.;
 RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
 RT grown at 303 K shows different orientations of the N and C lobes.";
 RL Acta Crystallogr. B 55:1805-1813(1999).
 CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: SECRETED.

```

CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AJ005203; CA06441.1; -.
DR PDB: 1CE2; 19-MAR-99.
DR INTERPRO: IPR001156; -.
DR PFAM: PF00405; transferrin; 2.
DR PRINTS: PS00422; TRANSFERRIN.
DR PROSITE: PS00205; TRANSFERRIN_1; 2.
DR PROSITE: PS00206; TRANSFERRIN_2; 2.
DR PROSITE: PS00207; TRANSFERRIN_3; 2.
KW Iron transport; Glycoprotein; Metal-binding; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 708 LACTOTRANSFERRIN.
FT DISULFID 28 64
FT DISULFID 38 55
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 179 202
FT DISULFID 189 200
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79
FT METAL 111 111 IRON 1.
FT METAL 211 211 IRON 1.
FT METAL 272 272 IRON 1.
FT METAL 414 414 IRON 2.
FT METAL 452 452 IRON 2.
FT METAL 545 545 IRON 2.
FT METAL 614 614 IRON 2.
FT BINDING 140 140 ANION (POTENTIAL).
FT BINDING 482 482 ANION (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 708 AA; 77729 MW; 08D2600AAB2F9ACD CRC64;
Query Match 5.3%; Score 91; DB 1; Length 708;
Best Local Similarity 24.9%; Pred. No. 4.3;
Matches 49; Conservativity 25; Mismatches 61; Indels 62; Gaps 10;

```

```

OY 219 -----SLKVPYALIPS 230
DB 255 RAPVDAFEKCHLAQVPS 271
RESULT 12
FBN1_BOVIN STANDARD; PRT; 2871 AA.
ID FBN1_BOVIN
AC P98133;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FIBRILLIN 1 PRECURSOR (MP340).
OS FBN1.
GN Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-SKIN;
RX MEDLINE=95137597; PubMed=7835900;
RA Tilstra D.J., Potter K.A., Byers P.H.;
RT "Sequence of the coding region of the bovine fibrillin cDNA and
RL localization to bovine chromosome 10.";
RN Genomics 23:480-485(1994).
[2]
RP PARTIAL SEQUENCE.
RX MEDLINE=96132851; PubMed=8557636;
RA Gibson M.A., Hatzinikolas G., Kumaratilake J.S., Sandberg L.B.,
RA Nicholl J.K., Sutherland G.R., Cleary E.G.;
RT "Further characterization of proteins associated with elastic fiber
RT microfibrils including the molecular cloning of MAGP-2 (MP25).";
RL J. Biol. Chem. 271:1096-1103(1996).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS.
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: L28748; AA74122.1; -.
DR HSSP: P35553; IAPU.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR001438; -.
DR INTERPRO: IPR001881; -.
DR INTERPRO: IPR002212; -.
DR PFAM: PF00008; EGF; 46.
DR PFAM: PF00683; TB; 9.
DR PRINTS: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_CA; 45.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.

```


FT	DOMAIN	246	287	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	288	329	EGF-LIKE 5, CALCIUM-BINDING.
FT	REPEAT	330	390	TGFBP 1.
FT	DOMAIN	392	446	PRO-RICH.
FT	DOMAIN	449	489	EGF-LIKE 6, NON-CALCIUM BINDING.
FT	DOMAIN	490	529	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	530	571	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	572	612	EGF-LIKE 9, CALCIUM-BINDING.
FT	DOMAIN	613	653	EGF-LIKE 10, CALCIUM-BINDING.
FT	REPEAT	654	722	TGFBP 2.
FT	DOMAIN	723	764	EGF-LIKE 11, CALCIUM-BINDING.
FT	DOMAIN	765	806	EGF-LIKE 12, CALCIUM-BINDING.
FT	DOMAIN	807	846	EGF-LIKE 13, CALCIUM-BINDING.
FT	DOMAIN	910	951	EGF-LIKE 14, CALCIUM-BINDING.
FT	REPEAT	952	1027	TGFBP 3.
FT	DOMAIN	1028	1069	EGF-LIKE 15, CALCIUM-BINDING.
FT	DOMAIN	1070	1112	EGF-LIKE 16, CALCIUM-BINDING.
FT	DOMAIN	1113	1154	EGF-LIKE 17, CALCIUM-BINDING.
FT	DOMAIN	1155	1196	EGF-LIKE 18, CALCIUM-BINDING.
FT	DOMAIN	1197	1237	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN	1238	1279	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN	1280	1321	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN	1322	1362	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN	1363	1403	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN	1404	1445	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN	1446	1486	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN	1487	1527	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT	1528	1605	TGFBP 4.
FT	DOMAIN	1606	1647	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1648	1688	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT	1689	1765	TGFBP 5.
FT	DOMAIN	1766	1807	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1808	1848	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1849	1890	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1891	1929	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1930	1972	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	1973	2012	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2013	2054	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT	2055	2126	TGFBP 6.
FT	DOMAIN	2127	2165	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2166	2205	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2206	2246	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT	2333	2401	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING.
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	271	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.

FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.
FT	DISULFID	1326	1339	BY SIMILARITY.
FT	DISULFID	1333	1348	BY SIMILARITY.
FT	DISULFID	1350	1361	BY SIMILARITY.
FT	DISULFID	1367	1380	BY SIMILARITY.
FT	DISULFID	1374	1389	BY SIMILARITY.
FT	DISULFID	1391	1402	BY SIMILARITY.
FT	DISULFID	1408	1420	BY SIMILARITY.
FT	DISULFID	1415	1429	BY SIMILARITY.
FT	DISULFID	1431	1444	BY SIMILARITY.
FT	DISULFID	1431	1444	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1456	1470	BY SIMILARITY.
FT	DISULFID	1472	1485	BY SIMILARITY.
FT	DISULFID	1491	1502	BY SIMILARITY.
FT	DISULFID	1497	1511	BY SIMILARITY.
FT	DISULFID	1513	1526	BY SIMILARITY.
FT	DISULFID	1610	1622	BY SIMILARITY.
FT	DISULFID	1617	1631	BY SIMILARITY.
FT	DISULFID	1633	1646	BY SIMILARITY.
FT	DISULFID	1652	1663	BY SIMILARITY.

Query Match

Best Local Similarity 5.28; Score 90.5; DB 1; Length 2871;

Matches 52; Conservative 34; Mismatches 98; Indels 77; Gaps 12;

QY	99	OPSLCHKFIASDDEGKILRYTONIDTLEQVAGIORILOCHGSFATASCLICKRYVDC	158
DB	227	QPHCCRRCFIPDITRGA-----CODVDCQAIPLG-----COG-----GNCINTVGSFECK	272
QY	159	AVRQDIFNQVAPRCPRADEPLAIKKEIYFGE-----NLPEQFHRAMKRYDKD	208
DB	273	CPAGHKRFNEVSKCE--DIDCSIT--PGICDGGECFTVSSYRCKCPGFTYSPDGR	327

OY 209 EVDLLIVIGSLKVRP---VALIPSSIPHEVPOILLINREPLLPHLFVDELLGDCD---- 260
DB 338 -----CIDVRPGCYCTALANGRCNSNLPPOSITKMOCC-----CDAGRC 365
OY 261 -----VIINELCH-RUGGYAKLCCNPVKLSITEKPPRPQKELYHLSRLP----- 306
DB 366 WSPGVTVAPEMCPIRATEDENFKLSVPMVIPERPGVPPPLDGPVPVOPVPPGPPGPOI 425
OY 307 ----TPLHISEDSSSPERTVP 323
DB 426 MTRPREVPEYPSRPREPVLP 446

RESULT 13
HCVG_OCTDO STANDARD. PRT: 2896 AA.
ID HCVG_OCTDO STANDARD. PRT: 2896 AA.
AC 061363. 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMOCYANIN G-TYPE, UNITS ODA TO ODG.
GN ODHCV.
OS Octopus dofileini (Giant octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Octopoda;
OC Incirrata; Octopodidae; Octopus.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-BRANCHIAL GLAND:
RX MEDLINE=98277150; PubMed=9614945;
RA Miller K.I., Cuff M.E., Lang W.F., Varga-Welsz P., Field K.G.,
RA van Holde K.E.;
RT "Sequence of the Octopus dofileini hemocyanin subunit: structural and
RT evolutionary implications";
RL J. Mol. Biol. 278:827-842(1998).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF SUBUNIT ODG.
RX MEDLINE=98277152; PubMed=9614947;
RA Cuff M.E., Miller K.I., van Holde K.E., Hendrickson W.A.,
RA van Holde K.E.;
RT "Crystal structure of a functional unit from Octopus hemocyanin";
RL J. Mol. Biol. 278:855-870(1998).
CC -1- FUNCTION: HEMOCYANIN ARE COPPER-CONTAINING OXYGEN CARRIERS
CC OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
CC ARTHROPODS.
CC -1- COPFACTOR: CONTAINS TWO COPPER ATOMS PER FUNCTIONAL UNIT.
CC -1- SUBUNIT: DECAMERS OF LARGE IDENTICAL SUBUNITS (350 KDA). EACH
CC CONTAINING 7 GLOBULAR OXYGEN-BINDING FUNCTIONAL UNITS: ODA, ODB,
CC ODC, ODD, ODE, ODF, AND ODG. DECAMER FORMATION REQUIRES THE
CC PRESENCE OF MAGNESIUM IONS.
CC -1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC -----
CC DR EMBL; AF020548; AAC39018.1; -
CC INTERPRO: IPR002227; -
DR PFAM: PF00264; tyrosinase; 8.
DR PROSITE: PS00497; TYROSINASE_1; 4.
DR PROSITE: PS00498; TYROSINASE_2; 6.
KW Respiratory protein; Oxygen transport; Copper; Glycoprotein;
KW Hemolymph; Duplication.
FT DOMAIN 1 419 ODA.
FT DOMAIN 420 834 ODB.
FT DOMAIN 835 1254 ODC.
FT DOMAIN 1255 1667 ODE.
FT DOMAIN 1668 2085 ODF.

FT DOMAIN 2086 2502 ODF.
FT DOMAIN 2503 2896 ODG.
FT METAL 41 60 COPPER A (BY SIMILARITY).
FT METAL 60 60 COPPER A (BY SIMILARITY).
FT METAL 69 69 COPPER A (BY SIMILARITY).
FT METAL 178 178 COPPER B (BY SIMILARITY).
FT METAL 182 182 COPPER B (BY SIMILARITY).
FT METAL 209 209 COPPER B (BY SIMILARITY).
FT METAL 460 460 COPPER A (BY SIMILARITY).
FT METAL 480 480 COPPER A (BY SIMILARITY).
FT METAL 489 489 COPPER A (BY SIMILARITY).
FT METAL 601 601 COPPER B (BY SIMILARITY).
FT METAL 605 605 COPPER B (BY SIMILARITY).
FT METAL 632 632 COPPER B (BY SIMILARITY).
FT METAL 875 875 COPPER A (BY SIMILARITY).
FT METAL 895 895 COPPER A (BY SIMILARITY).
FT METAL 904 904 COPPER A (BY SIMILARITY).
FT METAL 1013 1013 COPPER B (BY SIMILARITY).
FT METAL 1017 1017 COPPER B (BY SIMILARITY).
FT METAL 1044 1044 COPPER B (BY SIMILARITY).
FT METAL 1292 1292 COPPER A (BY SIMILARITY).
FT METAL 1312 1312 COPPER A (BY SIMILARITY).
FT METAL 1321 1321 COPPER A (BY SIMILARITY).
FT METAL 1425 1425 COPPER B (BY SIMILARITY).
FT METAL 1429 1429 COPPER B (BY SIMILARITY).
FT METAL 1456 1456 COPPER B (BY SIMILARITY).
FT METAL 1708 1708 COPPER A (BY SIMILARITY).
FT METAL 1728 1728 COPPER A (BY SIMILARITY).
FT METAL 1737 1737 COPPER A (BY SIMILARITY).
FT METAL 1849 1849 COPPER B (BY SIMILARITY).
FT METAL 1853 1853 COPPER B (BY SIMILARITY).
FT METAL 1880 1880 COPPER B (BY SIMILARITY).
FT METAL 2126 2126 COPPER A (BY SIMILARITY).
FT METAL 2144 2144 COPPER A (BY SIMILARITY).
FT METAL 2153 2153 COPPER A (BY SIMILARITY).
FT METAL 2262 2262 COPPER B (BY SIMILARITY).
FT METAL 2266 2266 COPPER B (BY SIMILARITY).
FT METAL 2293 2293 COPPER B (BY SIMILARITY).
FT METAL 2543 2543 COPPER A (BY SIMILARITY).
FT METAL 2562 2562 COPPER A (BY SIMILARITY).
FT METAL 2571 2571 COPPER A (BY SIMILARITY).
FT METAL 2671 2671 COPPER B (BY SIMILARITY).
FT METAL 2675 2675 COPPER B (BY SIMILARITY).
FT METAL 2702 2702 COPPER B (BY SIMILARITY).
FT METAL 47 57 BY SIMILARITY.
FT DISULFD 168 234 BY SIMILARITY.
FT DISULFD 321 333 BY SIMILARITY.
FT DISULFD 466 477 BY SIMILARITY.
FT DISULFD 591 657 BY SIMILARITY.
FT DISULFD 881 892 BY SIMILARITY.
FT DISULFD 1003 1070 BY SIMILARITY.
FT DISULFD 1298 1309 BY SIMILARITY.
FT DISULFD 1415 1482 BY SIMILARITY.
FT DISULFD 1571 1581 BY SIMILARITY.
FT DISULFD 1714 1725 BY SIMILARITY.
FT DISULFD 1839 1906 BY SIMILARITY.
FT DISULFD 1997 2003 BY SIMILARITY.
FT DISULFD 2131 2141 BY SIMILARITY.
FT DISULFD 2252 2319 BY SIMILARITY.
FT DISULFD 2406 2411 BY SIMILARITY.
FT DISULFD 2549 2559 BY SIMILARITY.
FT DISULFD 2661 2728 BY SIMILARITY.
FT DISULFD 2815 2821 BY SIMILARITY.
FT THIOETH 58 60 BY SIMILARITY.
FT THIOETH 478 480 BY SIMILARITY.
FT THIOETH 893 895 BY SIMILARITY.
FT THIOETH 1310 1312 BY SIMILARITY.
FT THIOETH 1726 1728 BY SIMILARITY.
FT THIOETH 2142 2144 BY SIMILARITY.
FT THIOETH 2560 2562 BY SIMILARITY.
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1496 1496 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1634 1634 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2055 2055 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2201 2201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2553 2553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2896 AA; 331917 MW; BELF35C8C987FBFC CRC64;

Query Match 5.1%; Score 89; DB 1; Length 2896;
 Best Local Similarity 18.9%; Pred. No. 36; Mismatches 115; Indels 96; Gaps 14;
 Matches 60; Conservative 46;

QY 14 RKDINTIEDA-VKLECKRIYLTGA-----GVSVSC-----GIPIF 50
 DB 1258 RKNLNLTDGMEGLRAKAFDMTDRYEELASFGHLPACPKNGSKVYCCIHGMPTF 1317
 QY 51 RSRBGITARL-----AVDPDLPPQAMF-----IEYFRK-----DPRP 85
 DB 1318 PHMHLVVALVENELLARGSGVAVPYDWQV---FDHLPALVAVRATYNSRTLLVEPNP 1374
 QY 86 FFKFAKEIYFGQFQPSLCHKFIASDKGKL-LRNYQNTIDLEOVAGIORILLOCHGSA 144
 DB 1375 FFK-----GKISFLNSETNRDPOELFG-NKYLEHTLEV 1408
 QY 145 TASCLCKYVDEAVNGDIFNQVPRCPRADEPLAIKPEIVFGENLPE----- 197
 DB 1409 LEQTFDCEFEVHEVGLNTHSWLGRDPSMSLDAADPIFFLHNSNDRIWAIMQE 1468
 QY 198 -QFRAMKYNKDEVDLLIVGSSLKVRVALIPS-----SIPHEVPOILINREPLPHLF 251
 DB 1469 LQRYRKLPYNEANCLPLL---NVMRPFSTNTANHDRLMTLHSAPNDVEDYQNV--LHY 1523
 QY 252 DVELLGDGVINELCH 268
 DB 1524 KYDTLSFYDLITQDLH 1540

RESULT 14
 SAP_MOUSE STANDARD; PRT; 557 AA.
 ID SAP_MOUSE
 AC 061207; 064219; 064006; 060861;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SULFATED GLYCOPROTEIN 1 PRECURSOR (SGP-1) (PROSAPOSIN).
 GN PSAP OR SGPI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=92272718; PubMed=1590788;
 RA Tsuda M., Sakiyama T., Endo H., Kitagawa T.;
 RT "The primary structure of mouse saposin.";
 RN Biochem. Biophys. Res. Commun. 184:1266-1272(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94272317; PubMed=8003952;
 RA Sprecher-Ley H., Orr-Urtreger A., Ional P., Horowitz M.;
 RT "Murine prosaposin: expression in the reproductive system of a gene
 RT implicated in human genetic disease.";
 RL Cell. Mol. Biol. 40:233-233(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96084310; PubMed=8565332;
 RA Cao Q.P., Crain M.R.;
 RT "Expression of SGP-1 mRNA in preimplantation mouse embryos.";
 RL Dev. Genet. 17:263-271(1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=LIVER;
 RA Zhao Q.Q., Hay N.N., Morales C.R.;
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- SIMILARITY: CONTAINS FOUR SAPOSIN B-TYPE DOMAINS AND TWO A-TYPE
 CC DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: S36200; AAB22175.1; -;
 DR EMBL: S71616; AAB31059.1; -;
 DR EMBL: U27340; AAB92567.1; -;
 DR EMBL: U57999; AAB02695.1; -;
 DR MGI: 97783; PSAP.
 KW Sulfatation; signal; glycoprotein; Duplication; Repeat.
 FT SIGNAL 1
 FT CHAIN 17
 FT DOMAIN 21
 FT DOMAIN 524
 FT CARBOHYD 80
 FT CARBOHYD 214
 FT CARBOHYD 334
 FT CARBOHYD 459
 FT CARBOHYD 83
 FT CARBOHYD 158
 FT CARBOHYD 160
 FT CARBOHYD 171
 FT CARBOHYD 244
 FT CARBOHYD 254
 FT CARBOHYD 255
 FT CARBOHYD 262
 FT CARBOHYD 307
 FT CARBOHYD 322
 FT CARBOHYD 349
 FT CARBOHYD 367
 FT CARBOHYD 370
 FT CARBOHYD 373
 FT CARBOHYD 391
 FT CARBOHYD 393
 FT CARBOHYD 406
 FT CARBOHYD 430
 FT CARBOHYD 445
 FT CARBOHYD 448
 SQ SEQUENCE 557 AA; 61422 MW; 134593E2049E35E CRC64;
 G -> D (IN REF. 3).
 L -> Q (IN REF. 2).
 I -> D (IN REF. 3).
 A -> T (IN REF. 3).
 R -> L (IN REF. 3).
 A -> R (IN REF. 2 AND 3).
 P -> R (IN REF. 2).
 C -> F (IN REF. 3).
 L -> P (IN REF. 4).
 MS -> SA (IN REF. 3).
 V -> L (IN REF. 2).
 M -> I (IN REF. 3).
 L -> W (IN REF. 2).
 MISSING (IN REF. 3).
 N -> D (IN REF. 2).
 E -> L (IN REF. 2).
 AL -> GV (IN REF. 1).
 G -> D (IN REF. 3).
 L -> Q (IN REF. 2).
 I -> D (IN REF. 3).
 A -> T (IN REF. 3).
 R -> L (IN REF. 3).
 A -> R (IN REF. 2 AND 3).
 P -> R (IN REF. 2).
 C -> F (IN REF. 3).
 L -> P (IN REF. 4).

Query Match 5.1%; Score 87.5; DB 1; Length 557;
 Best Local Similarity 22.7%; Pred. No. 6.3;
 Matches 55; Conservative 37; Mismatches 83; Indels 67; Gaps 14;

QY 114 GKLR-VYQW--IDTLEOVAGIORILLOQ---HSPFATSC--LICKY-KYDCAVNGDI 164
 DB 74 GNLLKNATQDEILHYEKT-----CEWIHSSLSASCEVVDSTLPYIIMIGEM 125
 QY 165 FN--QVPRPCRADEPLAIKPEIVFGENLPEQFRAKRYD-DEVDLLIVGSSLK 221
 DB 126 SNPEVCSALNLQOS-----LQEYLAEQNOKQLESKIKPEVDARVAPFMS 172
 QY 222 VRPALIPSSIPHEVPOILINREPLPHLHDV-----ELGDCDYI----- 262
 DB 173 NIPLLYPODHPRSPQ-----PRANEDVCODCKKLVSVDYQTAVKTNSSFYQGFVDH 224
 QY 263 INELCHRLGSEYAKLQCNPV-KISEITEKPRPRQKELVHSELPTPLHISEDSSPERT 321
 DB 225 VKEDCDRLGSEVSDICKNVDQYSEVC-----VQMLHMDQDQPKKEICVLAGECNEVKR 278
 QY 322 VP 323
 DB 279 VP 280

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:44:07 : Search time 51.38 Seconds
(without alignments)
114.285 Million cell updates/sec

Title: US-09-461-580A-12

Perfect score: 1732

Sequence: 1 VINILSEPPKKRRKDINTI.....PLHISEDSSSPERTVPQDS 327

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	5.6	708	1 US-08-145-681-4	Sequence 4, Appl1
2	97	5.6	708	1 US-08-453-703-4	Sequence 4, Appl1
3	97	5.6	708	2 US-08-456-106-4	Sequence 4, Appl1
4	97	5.6	708	3 US-08-456-108-4	Sequence 4, Appl1
5	86	5.0	329	3 US-09-108-020-48	Sequence 48, Appl1
6	82.5	4.8	579	1 US-08-448-196A-8	Sequence 8, Appl1
7	81	4.7	740	3 US-09-022-983-5	Sequence 5, Appl1
8	78.5	4.5	593	2 US-08-900-927-4	Sequence 4, Appl1
9	78.5	4.5	593	2 US-09-191-279-4	Sequence 4, Appl1
10	78.5	4.5	593	3 US-09-334-476-4	Sequence 4, Appl1
11	78	4.5	2414	1 US-08-327-536-2	Sequence 2, Appl1
12	78	4.5	2414	4 PCT-US95-04682-2	Sequence 2, Appl1
13	77.5	4.5	2366	1 US-08-480-604A-10	Sequence 10, Appl1
14	77.5	4.5	2366	2 US-08-405-496A-10	Sequence 10, Appl1
15	77	4.4	433	2 US-08-466-120-2	Sequence 2, Appl1
16	77	4.4	433	4 PCT-US94-07266-2	Sequence 2, Appl1
17	77	4.4	2555	3 US-09-058-489-36	Sequence 36, Appl1
18	76.5	4.4	1022	1 US-08-271-364A-8	Sequence 8, Appl1
19	76.5	4.4	1022	2 US-08-222-715B-27	Sequence 27, Appl1
20	76.5	4.4	2547	3 US-09-058-489-35	Sequence 35, Appl1
21	76	4.4	718	2 US-08-560-398-12	Sequence 12, Appl1
22	75	4.3	486	3 US-08-904-452-2	Sequence 2, Appl1
23	75	4.3	593	2 US-08-900-927-3	Sequence 3, Appl1
24	75	4.3	593	2 US-09-191-279-3	Sequence 3, Appl1
25	75	4.3	593	3 US-09-334-476-3	Sequence 3, Appl1
26	75	4.3	2366	1 US-08-157-005-2	Sequence 2, Appl1
27	74.5	4.3	484	3 US-09-263-023-4	Sequence 4, Appl1
28	74.5	4.3	523	1 US-08-100-247-2	Sequence 2, Appl1

29	74.5	4.3	523	1 US-08-483-146A-2	Sequence 2, Appl1
30	74.5	4.3	523	1 US-08-484-594A-2	Sequence 2, Appl1
31	74.5	4.3	893	1 US-07-977-434-4	Sequence 4, Appl1
32	74.5	4.3	893	1 US-08-458-819-4	Sequence 4, Appl1
33	74.5	4.3	893	4 PCT-US91-07035-4	Sequence 4, Appl1
34	74	4.3	274	2 US-07-857-224B-30	Sequence 30, Appl1
35	74	4.3	282	1 US-08-318-947A-19	Sequence 19, Appl1
36	74	4.3	282	2 US-08-795-503-19	Sequence 19, Appl1
37	74	4.3	297	3 US-08-874-347-22	Sequence 22, Appl1
38	74	4.3	297	3 US-09-093-522-22	Sequence 22, Appl1
39	73.5	4.2	523	1 US-08-232-513A-3	Sequence 3, Appl1
40	73	4.2	106	3 US-08-326-117B-6	Sequence 6, Appl1
41	73	4.2	106	3 US-08-982-129-6	Sequence 6, Appl1
42	73	4.2	157	1 US-08-494-577-4	Sequence 4, Appl1
43	73	4.2	157	1 US-08-494-577-5	Sequence 4, Appl1
44	73	4.2	157	2 US-08-795-868-4	Sequence 4, Appl1
45	73	4.2	157	2 US-08-795-868-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-145-681-4
; Sequence 4, Application US/08145681
; Patent No. 5571691
; GENERAL INFORMATION:
; APPLICANT: Connely, Orla M.
; APPLICANT: Heaton, Denis R.
; APPLICANT: O'Malley, Bert W.
; APPLICANT: May, Gregory S.
; TITLE OF INVENTION: Production of Recombinant Lactoferrin
; TITLE OF INVENTION: and Lactoferrin Polypeptides using cDNA Sequences in
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Baker & Botts, L.L.P.
; STREET: 910 Louisiana St
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,681
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcgregor, Martin L.
; REGISTRATION NUMBER: 29,329
; REFERENCE/DOCKET NUMBER: 19928-0125
; TELEPHONE: 713/229/1874
; TELEFAX: 713/229/1522
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGIN: SOURCE:
; ORGANISM: Bos taurus
; US-08-145-681-4
Query Match 5.6%; Score 97; DB 1; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.075;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

```
OY 81 KDPREFEAKFIYGGQPSLCHKFIALSDK-----EGK----- 115
Db 88 KDPYLRPAALAIYGTKEPQTHYAAVAVKGSNFOLDQLOGRSCHTGLRSAGWIIP 147
OY 116 --LLNNYTONIDITLBOVAGIQRILIOCHGSFATASCLIC---KYVDEAVRGDIENOVY 169
Db 148 MGLRPYLSTWTSLEPLQG-----AVAKFESASCVPICDRQAYPNLQOLCKGEGENQCA 201
OY 170 PRCPRCPADP-----LAIKPEIVFGENLPQGFHRAMKYDDEVDL 212
Db 202 -----CSSREPEFGYSGAFKCLDAGADVAFAVKEITVF--ENLPE-----KADRDQYEL 248
OY 213 LIVIGS-----SLKVPVALIPS 230
Db 249 LCLNNSRAPVDKFKCHLAQVPS 271
```

RESULT 2

US-08-453-703-4
Sequence 4, Application US/08453703
Patent No. 5766939

GENERAL INFORMATION:

APPLICANT: Conneely, Orla M.
APPLICANT: Headon, Denis R.
APPLICANT: O'Malley, Bert W.
APPLICANT: May, Gregory S.
TITLE OF INVENTION: Production of Recombinant Lactoferrin
TITLE OF INVENTION: and Lactoferrin Polypeptides Using cDNA Sequences in
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453.703
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/145,681
FILING DATE: October 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8206-024
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 708 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bos taurus
US-08-453-703-4

Query Match 5.6%; Score 97; DB 1; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.075;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

```
OY 81 KDPREFEAKFIYGGQPSLCHKFIALSDK-----EGK----- 115
Db 88 KDPYLRPAALAIYGTKEPQTHYAAVAVKGSNFOLDQLOGRSCHTGLRSAGWIIP 147
OY 116 --LLNNYTONIDITLBOVAGIQRILIOCHGSFATASCLIC---KYVDEAVRGDIENOVY 169
Db 148 MGLRPYLSTWTSLEPLQG-----AVAKFESASCVPICDRQAYPNLQOLCKGEGENQCA 201
OY 170 PRCPRCPADP-----LAIKPEIVFGENLPQGFHRAMKYDDEVDL 212
Db 202 -----CSSREPEFGYSGAFKCLDAGADVAFAVKEITVF--ENLPE-----KADRDQYEL 248
OY 213 LIVIGS-----SLKVPVALIPS 230
Db 249 LCLNNSRAPVDKFKCHLAQVPS 271
```

RESULT 3

US-08-456-106-4
Sequence 4, Application US/08456106
Patent No. 5849881

GENERAL INFORMATION:

APPLICANT: Conneely, Orla M.
APPLICANT: Headon, Denis R.
APPLICANT: O'Malley, Bert W.
APPLICANT: May, Gregory S.
TITLE OF INVENTION: Production of Recombinant Lactoferrin
TITLE OF INVENTION: and Lactoferrin Polypeptides Using cDNA Sequences in
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456.106
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/145,681
FILING DATE: October 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8206-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 708 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bos taurus
US-08-456-106-4

Query Match 5.6%; Score 97; DB 2; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.075;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDPPEFFKFKKEIYGGGQPSLCHKFIATLSDK-----EGK----- 115
DB 88 KDPKRLRPVAAEITGYESPQTHYAAVAVKGSNFOLDLOQGRKSCHTGLGNSAGNIIP 147
QY 116 --LLRNTYONIDTLEQVAGIQRILQCHGSFATASCLIC---KKVDCNAVGRDIFNQV 169
DB 148 MGLIRPLSLWTSLEPIQG-----AAKFFSASCVCIDRQAVPNICQLCKGSGENQCA 201
QY 170 PRCPRCAPDEP-----LAIMKEIYFEGENLPEQFHRANKYDKDEVLD 212
DB 202 -----CSSREPYFGYSAFKCLQDAGDAVAFKETVFE--ENLPE-----KADRDQYEL 248
QY 213 LIVIGS-----SLKVRPALIPS 230
DB 249 LCLNNSRAPVDARECHLAQVPS 271

RESULT 4
US-08-456-108-4
; Sequence 4, Application US/08456108
; Patent No. 6100054
; GENERAL INFORMATION:
; APPLICANT: Conneely, Orla M.
; APPLICANT: Headon, Denis R.
; APPLICANT: O'Malley, Bert W.
; TITLE OF INVENTION: Production of Recombinant Lactoferrin
; TITLE OF INVENTION: and Lactoferrin Polypeptides using cDNA Sequences in
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennine & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,108
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/145,681
; FILING DATE: October 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; US-08-456-108-4

Query Match 5.6%; Score 97; DB 3; Length 708;
Best Local Similarity 23.6%; Pred. No. 0.075;

Matches 48; Conservative 26; Mismatches 57; Indels 72; Gaps 9;

QY 81 KDPPEFFKFKKEIYGGGQPSLCHKFIATLSDK-----EGK----- 115
DB 88 KDPKRLRPVAAEITGYESPQTHYAAVAVKGSNFOLDLOQGRKSCHTGLGNSAGNIIP 147
QY 116 --LLRNTYONIDTLEQVAGIQRILQCHGSFATASCLIC---KKVDCNAVGRDIFNQV 169
DB 148 MGLIRPLSLWTSLEPIQG-----AAKFFSASCVCIDRQAVPNICQLCKGSGENQCA 201
QY 170 PRCPRCAPDEP-----LAIMKEIYFEGENLPEQFHRANKYDKDEVLD 212
DB 202 -----CSSREPYFGYSAFKCLQDAGDAVAFKETVFE--ENLPE-----KADRDQYEL 248
QY 213 LIVIGS-----SLKVRPALIPS 230
DB 249 LCLNNSRAPVDARECHLAQVPS 271

RESULT 5
US-09-108-020-48
; Sequence 48, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luehly, Michael H.
; APPLICANT: Mooney, Brian P.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: DMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 48
; LENGTH: 329
; TYPE: PRT
; ORGANISM: M. capricolm
; US-09-108-020-48

Query Match 5.0%; Score 86; DB 3; Length 329;
Best Local Similarity 20.1%; Pred. No. 0.37;
Matches 72; Conservative 45; Mismatches 133; Indels 108; Gaps 16;

QY 16 DINTEDAVKLLQCECKRIIVLTGAGVSYSCGIPFRSNDGIYALANDP-----DLRDP 70
DB 6 NIKAVTALDCAMQDRPNVIVFEGDVGTGGV--FRAQG---LAKFGNDRCFNATIS 59
QY 71 QAMF---DIEYFRKDPPEFFKFKKEIYGGGQPSLCHKFIATLSDKGLNRYONTIDTL 127
DB 60 EAMFAGVGLGMAANGMKFVLEMO---FEGGLAGLQNTFTNISMRRNTRGKTAPAWYIR 116
QY 128 EQVAGIQRILQCHGSFATASCLICKKVDCNAVGRDIFN-OVPRCPRCAPDEP---LA 182
DB 117 MPKGGGIGIALEHSE-----ALBAVYAHIPGVQIV--CRSTPYDTGLILAA 161
QY 183 IMKEIYFEGENLPEQFHRANKYDKDEVLDLIVIGSSIKY-----R 223
DB 162 IDSPDPVIVVE--PKTKYRAKQEPDDEHYIVPIGEGYKIOEGNDLVYVYGAOTVDCQK 219
QY 224 PVALIPSSIPH-----EVPQILINREPLPHLHFVDLLGDC 259
DB 220 AIALKETHRNATIDLIDRSIKPMDKKNVIESYKTKGRLLVHAEVAVSFSVAEII--- 276

```

QY 260 DVIINELCHRLGGELAKLCCNPVKLSLSETEK-----PPRQKELVHSEL 304
      : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 -ATVNEECF---EYIK-----APLSRCTGYDVITPFDREGGYFQVNBKPKVLVKKQEL 324

```

RESULT 6
NIS-08-AA8-

US-08-448-196A-8
Sequence 8, Application US/08448196A
Patent No. 5780594
GENERAL INFORMATION:
APPLICANT: CARTER, DANIEL C.
TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN FRAGMENTS
TITLE OF INVENTION: CONTAINING SPECIFIC BINDING REGIONS OF SERUM ALBUMIN OR
TITLE OF INVENTION: RELATED PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NASA
STREET: MARSHALL SPACE FLIGHT CENTER
CITY: HUNTSVILLE
STATE: ALABAMA
COUNTRY: USA
ZIP: 35812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,196A
FILING DATE: 23-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: BROAD JR., ROBERT L.
REGISTRATION NUMBER: 18,757
REFERENCE/DOCKET NUMBER: XX/MFS-28402-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 205-544-0021
TELEFAX: 205-544-0258
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-448-196A-8

[illegible]

RESULT 7

US-09-022-983-5
; Sequence 5, Application US/09022983
; Detest No. 5150731

GENERAL INFORMATION:

APPLICANT: Yang, Xiaoli
 APPLICANT: Khosravi-Far, Roya
 APPLICANT: Chang, Howard Y.
 APPLICANT: Baltimore, David
 TITLE OF INVENTION: DAXX, A NOVEL FAS-BINDING
 TITLE OF INVENTION: PROTEIN THAT ACTIVATES JNK AND APOPTOSIS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 STREET: 600 Atlantic Avenue
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/022,983
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/037,919
 FILING DATE: 12-FEB-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/051,753
 FILING DATE: 26-JUN-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Van Amsterdam, John R.
 REGISTRATION NUMBER: 40,212
 REFERENCE/DOCKET NUMBER: M0656/7036
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 740 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-09-022,983-5

Query Match	4.7%	Score 81	DB 3	Length 740
Best Local Similarity	22.1%	Pred. No. 4.3		
Matches 42	Conservative	22	Mismatches 64	Indels 62
				Gaps 7

QY	140	HGFATAS----	CLICKKYVDCAEVRGDI	FNQVVRPCRP	CPADEPLAI	IMKDEIV	EGGEN	194
		: : :					:	
Db	45	HGARSSSSGGKCC--	YKLENE----	KLFEELFC	KMOTADH----	DEVVFLYN	90	
		: : :						
QY	195	LPEQHRAMKXDKDEVD	LLIVGSLKVR	PVALI	ISSIHENP	QILINER	PHILAFVE	254
		: : :	: : :	:				
Db	91	ROQRHSHLFLASAE	CNLLSRVLS	RSRSPAKLY	-----			124
		: : :	: : :					
QY	255	LLGDGVILINCHRL	AGEYAKL	CNPVKLS	ITEETK	PPRQKELV	HLSELPTPLHISD	314
				: : :				
Db	125	-----VYINEL	CTVLRK	HA	SAKKRL	INLA	PAATTSNEPS-----	GNNPPT--HLISD
								167
QY	315	SSSPERTV	PQ	324				
		: : :						
Db	168	PTNAENTASQ	177					

RESULT 8

Db 803 VN-----SPL-----MPPGSGSHIHCPQLPQALHNSPSPVPSRT 839

```

RESULT 12
PCT-US95-04682-2
; Sequence 2, Application PC/US9504682
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
; NUMBER OF INVENTIONS: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurjin, Gagneblin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,536
; FILING DATE: 14-April-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Heine, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-308X999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04682-2

```

Query Match 4.5%; Score 78; DB 4; Length 2414;
 Best Local Similarity 19.4%; Pred. No. 51;
 Matches 58; Conservative 39; Mismatches 110; Indels 92; Gaps 13;

```

QY 64 FPDLPDQAFD-----IEYFRDPPRFKFAKEIYPGQFQPSLCHKFTALSDKEGKL 117
Db 592 FP--TPDPAALKDRMENLVAVARKVEDMYESANN--RAEYHLLAEKITYIQELEKEK 648
QY 118 RNTQNTDITLQVAGI-----QRIQCHGSFATASCLICKYVDEAARGD 163
Db 649 RTRLQKQNMJPNAGMVPVSNMPPNGQPQPGMTSNGPLP-----DPSMIRGS 697
QY 164 IFNOVVR-CPRCPADE--PLAIKMKPEIV-----FFGE-----NLPEOFHRAKY 205
Db 698 VPNQNMRRITPQSLNPGQMSMAQPIVPRQTPRLQHGGLADPAGALNPMGIGPRMQQ 757
QY 206 DKDEVLLIIV--GSSIKVRPVALIPSSIPHEVPOILINREPLPHLHFDVELLGDQVI 262
Db 758 PSNCGQFLPQTQFSPQGMNVTNIPLAPSSQAPVSAQOMSSSCP-----802
QY 263 INELCHNLGGEYAKLCNPVKLSITEKPPRQKEVLHLSLPTPLPHISDSSPEPT 321
Db 803 VN-----SPL-----MPPGSGSHIHCPQLPQALHNSPSPVPSRT 839

```

RESULT 13

US-08-480-604A-10
 ; Sequence 10, Application US/08480604A
 ; Patent No. 5736139

```

; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-10

```

Query Match 4.5%; Score 77.5; DB 1; Length 2366;
 Best Local Similarity 17.8%; Pred. No. 57;
 Matches 54; Conservative 50; Mismatches 125; Indels 75; Gaps 8;

```

QY 2 INIIS-----EPPRRKKRKDDINTEDAKILOECKIIVLGGAGVSGGIPDF 50
Db 826 INVISNIDTQIVERIEBAKMLTSDSINYIKDERKLE-----SIDALCDL 872
QY 51 RSRDGIYARLAVDPDLPDQAFDIEYFRDPPRF--FFKFAKEIYPGQFQPSLCHKFTA 108
Db 873 KQNMLEDSHRISFEDISFETDEGFSIRINKETGESIFVETKTLF--SEVANHTEELSK 931

```

```
Oy   LSDK-----BEKLRL-----VTYQNIIDPLEGVAGIOKHIILOCHSPAMASGLICKRYVDOEA 159
      |||:: :|| ::|| |::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||
Db    IKGTIFDTVGSKLVKKYNLDITTHEVNLTINAAFTQSILIEYNSKSELSLMSLYAMKVQOYA  991
Oy    160 -----VRGDIFNQVPVPCRPAPDEPLATMKPEIVFFGENTLEPFCRHR 201
Db    992 QLFSTGLNTITDAKVELSVTALDEFIDBLPTLSBCLPIINTIIDGVSLGAINKELSET 1051
Oy     202 AMKIKDKDEV-----LLIYTGSLLKYRPAVALIPSSITPHEVPOL 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1052 SDPLLROEIEAKIGIMAVNLTATAITISLSGIAGEGFILLVLPLAGIASGIPSIVNNEL 1111
Oy     241 INRE 244
      : : : :
Db    1112 VLRD 1115
```

```

RESULT 14
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338.
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-10

```

Matches	54;	Conservative	50;	Mismatches	125;	Indels	75;	Gaps	8;
Qy	2	INITLS-----	-----EPPKRRKRDINTIEDAVNLLDCKEKKIYLVGTAGVSCGIDPF	50					
Db	836	INVININIDTQIWEERIEEAKNLTSDSINIKDEFKLT-----	-----SISDALCDL	872					
Qy	51	RSRDSGIYARLAVDPFDLPDPOAMFIEYERKDRP- -PFKFAKEIYPOGQFOSLCHKFIA	108						
Db	873	KQONLEDSHFISETDEISFTDEGFSIRFLNKTGTGESIFETEKTIF-SEYANHITTEISK	931						
Qy	109	LSDK-----EGKLLR-----	-----NYTONIDTLEOVYAGIORILLOCHGSFPATASCLLCKRYVDDCEA	159					
Db	932	IKGTIFEDTVNGKLVAKKVNIDTTHEVNTLNAAFIQLSLEFYNSKRESLSLSTYAMKVQYVA	991						
Qy	160	-----VRGDIENOVVPRCDRCAPADEPLAIKKEIYFEGENLPDEQFHR	201						
Db	992	QLFTSTLNTITPDAKAVVELSTALDETIDLPTELSEGLPIITITIDGVSIGAAIKELSET	1051						
Qy	202	AMKYRKDEVD-----	-----LLIYIGSSLKRPVALLPSSIFPHEVPQL	240					
Db	1052	SDPLLROETIEAKIGIMAVNLATTTATTAITTSIGIASGFSILLVPLAGISAGIPSPKVNDEL	1111						
Qy	241	INRE 244							
Db	1112	VLRD 1115							

RESULT 15
 US-08-466-120--2
 Sequence 2, Application US/08466120
 Patent No. 5869284
 GENERAL INFORMATION:
 APPLICANT: CAO, ET AL.
 TITLE OF INVENTION: Retinoic Acid Receptor Epsilon
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ADDRESSEE: CECCHI, STEWART & OLSTEIN
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,120
 FILING DATE: June 6, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PC7/US94/07266
 FILING DATE: 24 JUN 94
 ATTORNEY/AGENT INFORMATION:
 NAME: FERRARO, GREGORY D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-354
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 433 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-08-466-120--2

Query Match	4.58; Score 77.5; DB 2; Length 2366;
Best Local Similarity	17.88; Pred. No. 57;

Query Match 4.48; Score 77; DB 2; Length 433;

Best Local Similarity 24.3%; Pred. No. 5.2;
Matches 33; Conservative 15; Mismatches 46; Indels 42; Gaps 6;

Matches	33;	Conservative	15;	Mismatches	46;	Indels	42;	Gaps	6;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

```
QY      224 PVALI-----PSIDHEVPQLINLEPLPHL-----HFDELLGGDDVII 263
          ||| | | | : | : | : | : | : |
Db      63 PTALLTRAEPSEPTETLRPPKRRKGPAKKMIGNELGVCODKASGFNYNLTSCGGCEFF 122
          ||| | | | : | : | : | : | : |
QY      264 NE-----LCRLRG--GEVAKLCCNPVKLSE---ITKEPPOPKELYHLSLPPTP 308
          ||| | | | : | : | : | : | : |
Db      123 RRSVIKAHHYICHSGCHCPMMTTYIRKRRCQECVLSEQIRLKLLKROEEVOYHATSLPR- 181
          ||| | | | : | : | : | : | : |
QY      309 LHISEDSSSPERTVPQ 324
          :||| : :||
Db      182 -----ASSPEQLLPQ 191
```

Db 63 P T A L L T R A E P S E P F E I R P P K R K K G A P K M L G N E L C S V C G D K A S G F H Y N V L S C E G C K E F F 122

QY 264 NE-----LCHRLG-----GEYAKLCNPVKLSE-----ITEKPPRQKELVHLSLPPTP 308

Db 123 RRSVIGAHYICHSGGHCPRNNTYMRKCGECLSEEDIRLKLKRGEEQVHATSLPPR- 181

QY 309 LHISSDSSPERTVPQ 324

```

Db      182 -----ASSPQILPQ 191

```

Search completed: February 16, 2001, 10:44:12
Job time: 154 sec

Job time: 154 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:49:15 ; Search time 220.15 Seconds

(without alignments)
226.005 Million cell updates/sec

Title: US-09-461-580A-12

Perfect score: 1732
Sequence: 1 VINILSEPPKRRKKRDINTI.....PLHISEDSSSPERTYPODS 327

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 936741 seqs, 152156132 residues

Total number of hits satisfying chosen parameters: 936741

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA:*

1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US07.COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US08.COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US081.COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US082.COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US083.COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US084.COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US085.COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US086.COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US087.COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US088.COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US089.COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US090.COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US091.COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US092.COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US093.COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US094.COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US095.COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US096.COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US097.COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US098.COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US099.COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US100.COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US101.COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US102.COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US103.COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US104.COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US105.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1732	100.0	327	US-09-461-580A-12	Sequence 12, Appl
2	1732	100.0	737	US-09-461-580A-1	Sequence 1, Appl
3	1732	100.0	737	US-09-461-580A-26	Sequence 26, Appl
4	1288	74.4	245	US-09-461-580A-4	Sequence 4, Appl
5	1021	58.9	823	US-09-619-049-576	Sequence 576, App

6	1021	58.9	823	22	US-60-167-217-10885	Sequence 10885, A
7	1021	58.9	823	22	US-60-171-627-915	Sequence 915, App
8	1021	58.9	823	22	US-60-173-464-8827	Sequence 8827, Ap
9	1021	58.9	823	22	US-60-191-637-10901	Sequence 10901, A
10	1021	58.9	823	22	US-60-191-681-8549	Sequence 8549, Ap
11	750	43.3	286	16	US-09-270-767-44940	Sequence 44940, A
12	682	39.4	128	18	US-09-461-580A-9	Sequence 9, Appl
13	643.5	37.2	339	16	US-09-248-796-19599	Sequence 19599, A
14	643.5	37.2	339	22	US-60-096-409-19599	Sequence 19599, A
15	568	32.8	313	16	US-09-248-796-20918	Sequence 20918, A
16	568	32.8	313	22	US-60-096-409-20918	Sequence 20918, A
17	567	32.7	336	18	US-09-461-580A-11	Sequence 11, Appl
18	556	32.1	106	18	US-09-461-580A-19	Sequence 19, Appl
19	548.5	31.7	389	22	US-60-169-629-517	Sequence 517, App
20	548.5	31.7	389	22	US-60-187-470-517	Sequence 517, App
21	535.5	30.9	267	18	US-09-461-580A-3	Sequence 3, Appl
22	520	30.0	355	22	US-60-167-217-10557	Sequence 10557, A
23	520	30.0	355	22	US-60-173-464-8533	Sequence 8533, Ap
24	517	29.8	272	18	US-09-461-580A-2	Sequence 2, Appl
25	489.5	28.3	284	16	US-09-248-796-19606	Sequence 19606, A
26	489.5	28.3	284	22	US-60-096-409-19606	Sequence 19606, A
27	484	27.9	381	22	US-60-117-905-57	Sequence 57, Appl
28	479	27.7	257	22	US-60-095-827-17	Sequence 17, Appl
29	454.5	26.2	358	18	US-09-417-507-39938	Sequence 39938, A
30	441.5	25.5	230	1	PCT-US00-02237-9	Sequence 9, Appl
31	441.5	25.5	230	22	US-60-117-905-9	Sequence 9, Appl
32	406	23.4	183	18	US-09-417-507-28899	Sequence 28899, A
33	367.5	21.2	232	18	US-09-461-580A-35	Sequence 35, Appl
34	312.5	18.0	107	18	US-09-461-580A-20	Sequence 20, Appl
35	309.5	17.9	488	16	US-09-248-796-19601	Sequence 19601, A
36	309.5	17.9	488	22	US-60-096-409-19601	Sequence 19601, A
37	304	17.6	188	22	US-09-461-580A-10	Sequence 10, Appl
38	295	17.0	245	18	US-60-117-905-70	Sequence 70, Appl
39	279.5	16.1	248	18	US-09-450-966-6406	Sequence 6406, Ap
40	278.5	16.1	270	16	US-09-252-991A-19648	Sequence 19648, A
41	267.5	15.4	223	19	PCT-US99-29950-188	Sequence 188, App
42	267.5	15.4	223	19	US-09-591-316-189	Sequence 189, App
43	266.5	15.4	107	18	US-09-461-580A-16	Sequence 16, Appl
44	265	15.3	233	20	US-09-611-529-6441	Sequence 6441, Ap
45	262	15.1	310	1	PCT-US00-02237-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-461-580A-12
: Sequence 12, Application US/09461580A
: GENERAL INFORMATION:
: APPLICANT: Guarente, Leonard
: APPLICANT: Imai, Shin-ichiro
: APPLICANT: Armstrong, Christopher
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FastSeq for Windows Version 4.0
: SFO ID NO 12
: LENGTH: 327
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-461-580A-12

Query Match 100.0%; Score 1732; DB 18; Length 327;
Best Local Similarity 100.0%; Pred. No. 2,4e-171;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VINILSEPPKRRKKRDINTIEDAVKLLQECKIIVLGAGVSCGIPDFSRDGIYARL 60
|||||

```

Db      1 VINIISPPKRRKKRDINTIEDAVKLLQECKKIIVLTGAGVSVSGGIPDRSRDGIYARL 60
Qy      61 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 120
        |||||||
Db      61 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 120
Qy      121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 180
        |||||||
Db      121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 180
Qy      181 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 240
        |||||||
Db      181 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 240
Qy      241 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 300
        |||||||
Db      241 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 300
Qy      301 LSELPTPLHISEDSSSPERTVPODSS 327
        |||||||
Db      301 LSELPTPLHISEDSSSPERTVPODSS 327

```

```

RESULT 2
US-09-461-580A-1
; Sequence 1, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imal, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-1

```

```

Query Match      100.0%; Score 1732; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 8.3e-171;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VINIISPPKRRKKRDINTIEDAVKLLQECKKIIVLTGAGVSVSGGIPDRSRDGIYARL 60
        |||||||
Db      216 VINIISPPKRRKKRDINTIEDAVKLLQECKKIIVLTGAGVSVSGGIPDRSRDGIYARL 275
Qy      61 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 120
        |||||||
Db      276 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 335
Qy      121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 180
        |||||||
Db      336 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 395
Qy      181 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 240
        |||||||
Db      396 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 455
Qy      241 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 300
        |||||||
Db      456 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 515
Qy      301 LSELPTPLHISEDSSSPERTVPODSS 327
        |||||||
Db      516 LSELPTPLHISEDSSSPERTVPODSS 542

```

```

RESULT 3
US-09-461-580A-26
; Sequence 26, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imal, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-26

```

```

Query Match      100.0%; Score 1732; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 8.3e-171;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VINIISPPKRRKKRDINTIEDAVKLLQECKKIIVLTGAGVSVSGGIPDRSRDGIYARL 60
        |||||||
Db      216 VINIISPPKRRKKRDINTIEDAVKLLQECKKIIVLTGAGVSVSGGIPDRSRDGIYARL 275
Qy      61 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 120
        |||||||
Db      276 AVDFPDLDPQAMFDIEYFRKDPREFPKAKIYPGQFQPSLCHKFIALSDKEGKLLRNY 335
Qy      121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 180
        |||||||
Db      336 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVKDCEAVRGDIFNOVVRPCRPCPADBP 395
Qy      181 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 240
        |||||||
Db      396 LAIMPEIYFEGENLPEQFHRAMKYDKDEVLLIYIGSSLKVRPAVALLPSSIPHEVPQIL 455
Qy      241 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 300
        |||||||
Db      456 INREPLPHLHFDVELLGGCDVYINELCHRLGGEYAKLCNPKVLSIETEKPPRPQKELVH 515
Qy      301 LSELPTPLHISEDSSSPERTVPODSS 327
        |||||||
Db      516 LSELPTPLHISEDSSSPERTVPODSS 542

```

```

RESULT 4
US-09-461-580A-4
; Sequence 4, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imal, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-4

```


Query Match 74.4%; Score 1288; DB 18; Length 245;
Best Local Similarity 100.0%; Pred. No. 2, 4e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VINLSEPPKRRKRDINTEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 60
DB 1 VINLSEPPKRRKRDINTEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 60
QY 61 AVFPPDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 120
DB 61 AVFPPDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 120
QY 121 TONIDTLEQVAGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPADPEP 180
DB 121 TONIDTLEQVAGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPADPEP 180
QY 181 LAIMKPEIVEFGENLPQGFHRAKMYDKDEVLLIVIGSSLKVRVALIPSSIPHEVQIL 240
DB 181 LAIMKPEIVEFGENLPQGFHRAKMYDKDEVLLIVIGSSLKVRVALIPSSIPHEVQIL 240
QY 241 INREP 245
DB 241 INREP 245

RESULT 5

US-09-619-049-576
; Sequence 576, Application US/09619049
; GENERAL INFORMATION:
; APPLICANT: YANDELL, MARK
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSECTICIDAL
; TITLE OF INVENTION: TARGETS
; FILE REFERENCE: CLO00735
; CURRENT APPLICATION NUMBER: US/09/619, 049
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/171,590
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/171,627
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/175,763
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/175,685
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/186,663
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 60/187,241
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 1533
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 576
; LENGTH: 823
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-619-049-576

Query Match 58.9%; Score 1021; DB 20; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;

QY 3 NILSEPPKRRKRDINTEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 62
DB 194 HLNLEPRKRRKASVNTFEDVLSVKKSKITIVLTGAGVSVSCGIPDRSRDGIYARL 253
QY 63 DFDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 122
DB 254 DFDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 313
QY 123 NIDTLEQVAGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPD----- 176
DB 123 NIDTLEQVAGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPD----- 176

DB 314 NIDTLEQVAGIORILQCHGSFATASCTCRKCNADLRADIFQRIIPVCPQPNKEQS 373
QY 177 -----ADEPL-----AIMKPEIVEFGENLPQGFHRAKMYDKDEVLLIVIGSSLKVR 223
DB 374 VDAVAVATEEELRQLVENGIMKPDIVFGEGLPDEYHTVMTDQVDDLIVIGSSLKVR 433
QY 224 PVALIPSSIPHEVQILINREPLPHLHFDVLLDQCVIINELCHRLGGE---YAKLCN 280
DB 434 PVALIPSSIPHEVQILINREPLPHLHFDVLLDQCVIINELCHRLGGE---YAKLCN 280
QY 281 PVKLSITEKPPRQKELVHSELPPPLHISEDSSPERTVQDS 326
DB 494 ESVLTESKELMP-PEHSHNLH---HLHHRHCSSESEKQSDLT 535

RESULT 6

US-60-167-217-10885
; Sequence 10885, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00152
; CURRENT APPLICATION NUMBER: US/60/167, 217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10885
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-10885

Query Match 58.9%; Score 1021; DB 22; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;

QY 3 NILSEPPKRRKRDINTEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 62
DB 194 HLNLEPRKRRKASVNTFEDVLSVKKSKITIVLTGAGVSVSCGIPDRSRDGIYARL 253
QY 63 DFDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 122
DB 254 DFDLPDQAMFDIEYFRKDRPFPRFKAKETIPGQFOPSLCHKFTALSKEKLLRNY 313
QY 123 NIDTLEQVAGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPD----- 176
DB 314 NIDTLEQVAGIORILQCHGSFATASCTCRKCNADLRADIFQRIIPVCPQPNKEQS 373
QY 177 -----ADEPL-----AIMKPEIVEFGENLPQGFHRAKMYDKDEVLLIVIGSSLKVR 223
DB 374 VDAVAVATEEELRQLVENGIMKPDIVFGEGLPDEYHTVMTDQVDDLIVIGSSLKVR 433
QY 224 PVALIPSSIPHEVQILINREPLPHLHFDVLLDQCVIINELCHRLGGE---YAKLCN 280
DB 434 PVALIPSSIPHEVQILINREPLPHLHFDVLLDQCVIINELCHRLGGE---YAKLCN 280
QY 281 PVKLSITEKPPRQKELVHSELPPPLHISEDSSPERTVQDS 326
DB 494 ESVLTESKELMP-PEHSHNLH---HLHHRHCSSESEKQSDLT 535

RESULT 7
US-60-171-627-915
; Sequence 915, Application US/60171627
; GENERAL INFORMATION:
; APPLICANT: Yandell, Mark
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; FILE REFERENCE: CLO00177

```

; CURRENT APPLICATION NUMBER: US/60/171,627
; CURRENT FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 823
; TYPE: PRP
; ORGANISM: Drosophila
US-60-171-627-915
```

```

Query Match          58.9%; Score 1021; DB 22; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;
```

```

OY 3 NLTSEPPKRRKKDINTTIDAVKLLQECKIIVLTGAGVSGCIPDFRSRGITARLAV 62
DB 194 HLTNEPKRRNKLASVNTFEDVLSVKSSQIIVLTGAGVSGCIPDFRSTNGIARLH 253
OY 63 DEPDLPDPQAMFDIEYFRKDPFRPFKAKEIYPGOPSLCHKFTALSDKEGKLLRNTQ 122
DB 254 DEPDLPDPQAMFDINFKDPRPFYKFAREIYPGEPSPCHRFIKMETKGLLRNTQ 313
OY 123 NIDTLEOVAGIQRILQCHGSFATASCLICKYKVCDAVNGDIFNOVPRPCRP----- 176
DB 314 NIDTLERVAGIQRVIECHGSFSTASCTCRFCNMADALRADIFAQRIPVPCOPCKEES 373
OY 177 -----ADEPL-----AIKPEIYFGEENLPEOFHRAKMDKDEVDLLIYIGSSLKYR 223
DB 374 VASAVAVTEELERQLVENGIMKPDIVFEGGLPDEHYTMATDKDCDILLIYIGSSLKYR 433
OY 224 PVALIPSSIPHEVPQILINREPLPHLHFDVELLDGCDVIINELCHRLGE---YAKLCN 280
DB 434 PVAHIPSSIPATVPQILINREQLHHLKFDVELLDGSDVIINQICHLRSLNDQWQLCCD 493
OY 281 PVKLSITEKPPRPQKELVHLSLPTPLHISEDSSSPERTVPQDS 326
DB 494 ESVLTESKELMP-PESHNNHLH---HLHHRHCSESESRQSLDT 535
```

```

RESULT 8
US-60-173-464-8827
; Sequence 8827, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8827
; LENGTH: 823
; TYPE: PRP
; ORGANISM: Drosophila
US-60-173-464-8827
```

```

Query Match          58.9%; Score 1021; DB 22; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;
```

```

OY 3 NLTSEPPKRRKKDINTTIDAVKLLQECKIIVLTGAGVSGCIPDFRSRGITARLAV 62
DB 194 HLTNEPKRRNKLASVNTFEDVLSVKSSQIIVLTGAGVSGCIPDFRSTNGIARLH 253
OY 63 DEPDLPDPQAMFDIEYFRKDPFRPFKAKEIYPGOPSLCHKFTALSDKEGKLLRNTQ 122
DB 254 DEPDLPDPQAMFDINFKDPRPFYKFAREIYPGEPSPCHRFIKMETKGLLRNTQ 313
OY 123 NIDTLEOVAGIQRILQCHGSFATASCLICKYKVCDAVNGDIFNOVPRPCRP----- 176
```

```

DB 314 NIDTLERVAGIQRVIECHGSFSTASCTCRFCNMADALRADIFAQRIPVPCOPCKEES 373
OY 177 -----ADEPL-----AIKPEIYFGEENLPEOFHRAKMDKDEVDLLIYIGSSLKYR 223
DB 374 VASAVAVTEELERQLVENGIMKPDIVFEGGLPDEHYTMATDKDCDILLIYIGSSLKYR 433
OY 224 PVALIPSSIPHEVPQILINREPLPHLHFDVELLDGCDVIINELCHRLGE---YAKLCN 280
DB 434 PVAHIPSSIPATVPQILINREQLHHLKFDVELLDGSDVIINQICHLRSLNDQWQLCCD 493
OY 281 PVKLSITEKPPRPQKELVHLSLPTPLHISEDSSSPERTVPQDS 326
DB 494 ESVLTESKELMP-PESHNNHLH---HLHHRHCSESESRQSLDT 535
```

```

RESULT 9
US-60-191-637-10901
; Sequence 10901, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Ventier, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: CLO00392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10901
; LENGTH: 823
; TYPE: PRP
; ORGANISM: DROSOPHILA
US-60-191-637-10901
```

```

Query Match          58.9%; Score 1021; DB 22; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;
```

```

OY 3 NLTSEPPKRRKKDINTTIDAVKLLQECKIIVLTGAGVSGCIPDFRSRGITARLAV 62
DB 194 HLTNEPKRRNKLASVNTFEDVLSVKSSQIIVLTGAGVSGCIPDFRSTNGIARLH 253
OY 63 DEPDLPDPQAMFDIEYFRKDPFRPFKAKEIYPGOPSLCHKFTALSDKEGKLLRNTQ 122
DB 254 DEPDLPDPQAMFDINFKDPRPFYKFAREIYPGEPSPCHRFIKMETKGLLRNTQ 313
OY 123 NIDTLEOVAGIQRILQCHGSFATASCLICKYKVCDAVNGDIFNOVPRPCRP----- 176
DB 314 NIDTLERVAGIQRVIECHGSFSTASCTCRFCNMADALRADIFAQRIPVPCOPCKEES 373
OY 177 -----ADEPL-----AIKPEIYFGEENLPEOFHRAKMDKDEVDLLIYIGSSLKYR 223
DB 374 VASAVAVTEELERQLVENGIMKPDIVFEGGLPDEHYTMATDKDCDILLIYIGSSLKYR 433
OY 224 PVALIPSSIPHEVPQILINREPLPHLHFDVELLDGCDVIINELCHRLGE---YAKLCN 280
DB 434 PVAHIPSSIPATVPQILINREQLHHLKFDVELLDGSDVIINQICHLRSLNDQWQLCCD 493
OY 281 PVKLSITEKPPRPQKELVHLSLPTPLHISEDSSSPERTVPQDS 326
DB 494 ESVLTESKELMP-PESHNNHLH---HLHHRHCSESESRQSLDT 535
```

```

RESULT 10
US-60-191-681-8549
; Sequence 8549, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
```

FILE REFERENCE: C1000390
CURRENT APPLICATION NUMBER: US/60/191,681
CURRENT FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 30973
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8549
LENGTH: 823
TYPE: PR
ORGANISM: DROSOPHILA
US-60-191-681-8549

Query Match 58.9%; Score 1021; DB 22; Length 823;
Best Local Similarity 58.7%; Pred. No. 9e-97;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;

QY 3 NILSEPKRRKRDINTIEAVKILQCKKITVLTGAGVSVSCIPPRSRDGIYARLAV 62
DB 194 HLLNEPRRKLASVNTFEDVISLVKKSOKIYLTGAGVSVSCIPPRSRDGIYARLAV 253
QY 63 DFDLPPOAMFDIEYRKPDRPFKFAKEIYGPQSPICHRKFTALSDKRGKILRYTQ 122
DB 254 DFDLPPOAMFDINTFKRPRPFKFAKEIYGPQSPICHRKFTALSDKRGKILRYTQ 313
QY 123 NIDLEQVAGIORTLQCHGSFATASCLICKYKVCCEAVRGDIFENQVPRCP-- 176
DB 314 NIDLEQVAGIORTLQCHGSFATASCLICKYKVCCEAVRGDIFENQVPRCP-- 373
QY 177 -----ADEPL-----AIMKPEIVFGENLPQFHRAMKYDKDEVLLIVIGSSILKVR 223
DB 374 VDAASVATEEELRLQVLENGIMKPDIVFGEGLPDEHYTAVTDKVDCLLVIGSSILKVR 433
QY 224 PVALLPESIPHEVPOILLINREPLPHLHFDVLLGDCDVIINELCHRGGE---YAKICCN 280
DB 434 PVAHIESIPATYVPOILLINREPLPHLHFDVLLGDCDVIINELCHRGGE---YAKICCN 280
QY 281 PVKLSIEITERPPRQKELVHSELPPYPLHISEDSSPERTVPODS 326
DB 494 ESVLTESKELMP--PEHSNHHLH---HLHHRHSSSESENGSOLDT 535

RESULT 11
US-09-270-767-44940
Sequence 44940, Application US/09270767
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 44940
LENGTH: 286
TYPE: PR
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44940

Query Match 43.3%; Score 750; DB 16; Length 286;
Best Local Similarity 54.4%; Pred. No. 2.8e-69;
Matches 149; Conservative 27; Mismatches 30; Indels 68; Gaps 3;

QY 41 VSVSGIPDRSRDGIYARLAVDFDLPDPOAMFDIEYRKPDRPFKFAKEIYGPQSP 100
DB 1 VSVSGIPDRSRDGIYARLAVDFDLPDPOAMFDINTFKRPRPFKFAKEIYGPQSP 60
QY 101 SLCHKFTALSDKRGKILRYTQNTIDLEQVAGIORTLQCHGSFATASCLICKYKVCCEAV 160
DB 61 SPCHRFKMETKGLRLRYTQNTIDLEQVAGIORTLQCHGSFATASCLICKYKVCCEAV 120

QY 161 RGDIFNQVPRCP--ADEPL-----AIMKPEIVFGE----- 193
DB 121 RADIFAQRIYPCQCKQNKESVDASVAVTEELRLQVLENGIMKPDIVFGEGRCHRYTD 180
QY 194 -----NPEQFHRAMKYDKDEVLL 212
DB 181 LLYIINSLGKINARISNTIKRSYLLRDKLKIQSNAIYHIGLPDEHYTAVTDKVDCL 240
QY 213 LIVIGSSILKVRVALIPSPHEVPOILLINREPL 246
DB 241 LIVIGSSILKVRVAHIPSIPATYVPOILLINREQL 274

RESULT 12
US-09-461-580a-9
Sequence 9, Application US/09461580A
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard
APPLICANT: Imai, Shin-Ichiro
APPLICANT: Armstrong, Christopher
TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
FILE REFERENCE: 0050:1618-000
CURRENT APPLICATION NUMBER: US/09/461,580A
CURRENT FILING DATE: 1999-12-15
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 128
TYPE: PR
ORGANISM: Mus musculus
US-09-461-580a-9

Query Match 39.4%; Score 682; DB 18; Length 128;
Best Local Similarity 100.0%; Pred. No. 9.7e-63;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 IVLTGAGVSVSCIPDRSRDGIYARLAVDFDLPDPOAMFDIEYRKPDRPFKFAKEI 93
DB 1 IVLTGAGVSVSCIPDRSRDGIYARLAVDFDLPDPOAMFDIEYRKPDRPFKFAKEI 60
QY 94 YPGQFQPSLCHKFTALSDKRGKILRYTQNTIDLEQVAGIORTLQCHGSFATASCLICKY 153
DB 61 YPGQFQPSLCHKFTALSDKRGKILRYTQNTIDLEQVAGIORTLQCHGSFATASCLICKY 120
QY 154 KVDCAVR 161
DB 121 KVDCAVR 128

RESULT 13
US-09-248-796-19599
Sequence 19599, Application US/09248796
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796
CURRENT FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 28206
SEQ ID NO 19599
LENGTH: 339
TYPE: PR
ORGANISM: Candida albicans
US-09-248-796-19599

Query Match 37.2%; Score 643.5; DB 16; Length 339;
Best Local Similarity 45.2%; Pred. No. 4.4e-58;
Matches 136; Conservative 38; Mismatches 80; Indels 47; Gaps 5;

```

QY 11 RRRKRDINTIEDAVKLLQECKKIIVLTGAGVSVCIGIPDFRSRDGIYARLAVDPDLDP 70
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 13 RRRREDITINAKLKIENSKNIMVITGISTSLGIDPFRSSQGFYS--MIQHLGLSDP 70
QY 71 CAMFIEYFRKPRPFKFAKEIYPGQFOPSLCHKFIASDKGKLLRNTYONIDPLEOV 130
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 71 QEVFLDLFLNDPNIFYSTAHMILPPNHYSPLSHSFIKLQDKNTLLRNTYONIDLESY 130
QY 131 AGI--ORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPC----- 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 AGIHKENLQCHGSFATASCLITCGYKVDGEIIFPEIKNKEIYCFKCEVKSILKKGK 190
QY 176 -----PAD-----EPLAIKKEIYFEGENLPEQFHRAMKYDKDE 209
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 191 TKSRSKKKKKKKKKKPYDDEEEEGEYFHSFSGVMKPDITFEQDLPEMFKAINDINK 250
QY 210 VDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLPHLAFDVELLGDCC---VIINE 265
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 251 VDLVAVIGTSLKVAIVADIVGKIPHEIPQILNKDPINHCOPDVSILGYCDVASYIANE 310
QY 266 L 266
DB 311 L 311

RESULT 14
US-60-096-409-19599
; Sequence 19599, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19599
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-19599

```

Query Match 37.2%; Score 643.5; DB 22; Length 339;
 Best Local Similarity 45.2%; Pred. No. 4,4e-58;
 Matches 136; Conservative 38; Mismatches 80; Indels 47; Gaps 5;

```

QY 11 RRRKRDINTIEDAVKLLQECKKIIVLTGAGVSVCIGIPDFRSRDGIYARLAVDPDLDP 70
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 13 RRRREDITINAKLKIENSKNIMVITGISTSLGIDPFRSSQGFYS--MIQHLGLSDP 70
QY 71 CAMFIEYFRKPRPFKFAKEIYPGQFOPSLCHKFIASDKGKLLRNTYONIDPLEOV 130
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 71 QEVFLDLFLNDPNIFYSTAHMILPPNHYSPLSHSFIKLQDKNTLLRNTYONIDLESY 130
QY 131 AGI--ORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPC----- 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 AGIHKENLQCHGSFATASCLITCGYKVDGEIIFPEIKNKEIYCFKCEVKSILKKGK 190
QY 176 -----PAD-----EPLAIKKEIYFEGENLPEQFHRAMKYDKDE 209
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 191 TKSRSKKKKKKKKKKPYDDEEEEGEYFHSFSGVMKPDITFEQDLPEMFKAINDINK 250
QY 210 VDLIVIGSSLKVRVALIPSSIPHEVPOILINREPLPHLAFDVELLGDCC---VIINE 265
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 251 VDLVAVIGTSLKVAIVADIVGKIPHEIPQILNKDPINHCOPDVSILGYCDVASYIANE 310
QY 266 L 266
DB 311 L 311

```

```

RESULT 15
US-09-248-796-20918
; Sequence 20918, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20918
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20918

```

Query Match 32.8%; Score 568; DB 16; Length 313;
 Best Local Similarity 43.1%; Pred. No. 2.8e-50;
 Matches 119; Conservative 53; Mismatches 80; Indels 24; Gaps 4;

```

QY 14 RKDINT-----IEDAVKLLQECKKIIVLTGAGVSVCIGIPDFRSRDGIYA 58
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 RQDMSITTLKRTINMTPEPKYKLPDLISDSKAKIMVYTGAGISTSLGIDPFRSEKGLYN 60
QY 59 RLAVDFPDLDPQAMFIEYFRKDRPFKFAKEIYPGQFOPSLCHKFIASDKGKLLR 118
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 QLSK--LNLSDPOKVFDTQTFKREGLTYTIAHLVLPDQKFSLHAFKLQDKHKLRL 118
QY 119 NTONIDPLEQVAG--IORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRCPC- 175
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 NTONIDNLEQAGAKLEKLVGCHGSFAKACVSCQGFAGEKINYNIIRKQVPCATCW 178
QY 176 ----PADEPLAIKPEIYFEGENLPEQFHRAMKYKDEVDLIVIGSSLKVRVALIPSS 231
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 179 KNTQAPRIHFGAIKPTIYFEGEDLPERFHTLMDKLOQIDFLVIGTSLKVEPVASIER 238
QY 232 IPHEVPOILINREPLPHLAFDVELLGDCCDVITINELC 267
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 239 VPKVYKILINKDPIPNNGFNILQILGLDCCDVVSYLC 274

```

Search completed: February 16, 2001, 10:49:17
 Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:43:20 ; Search time 76.21 Seconds
(Without alignments)
47.560 Million cell updates/sec

Title: US-09-461-580A-19
Perfect score: 556
Sequence: 1 IIVLTGAGVSVSCGIPDRS.....NTQNTDLEQVAGIQRIILQ 106

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Genseq-36: *
1: /SIDSL/gcgdata/genseq/genseqp/AA1980.DAT: *
2: /SIDSL/gcgdata/genseq/genseqp/AA1981.DAT: *
3: /SIDSL/gcgdata/genseq/genseqp/AA1982.DAT: *
4: /SIDSL/gcgdata/genseq/genseqp/AA1983.DAT: *
5: /SIDSL/gcgdata/genseq/genseqp/AA1984.DAT: *
6: /SIDSL/gcgdata/genseq/genseqp/AA1985.DAT: *
7: /SIDSL/gcgdata/genseq/genseqp/AA1986.DAT: *
8: /SIDSL/gcgdata/genseq/genseqp/AA1987.DAT: *
9: /SIDSL/gcgdata/genseq/genseqp/AA1988.DAT: *
10: /SIDSL/gcgdata/genseq/genseqp/AA1989.DAT: *
11: /SIDSL/gcgdata/genseq/genseqp/AA1990.DAT: *
12: /SIDSL/gcgdata/genseq/genseqp/AA1991.DAT: *
13: /SIDSL/gcgdata/genseq/genseqp/AA1992.DAT: *
14: /SIDSL/gcgdata/genseq/genseqp/AA1993.DAT: *
15: /SIDSL/gcgdata/genseq/genseqp/AA1994.DAT: *
16: /SIDSL/gcgdata/genseq/genseqp/AA1995.DAT: *
17: /SIDSL/gcgdata/genseq/genseqp/AA1996.DAT: *
18: /SIDSL/gcgdata/genseq/genseqp/AA1997.DAT: *
19: /SIDSL/gcgdata/genseq/genseqp/AA1998.DAT: *
20: /SIDSL/gcgdata/genseq/genseqp/AA1999.DAT: *
21: /SIDSL/gcgdata/genseq/genseqp/AA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	301.5	54.2	212	21	V48540	Human breast tumor
2	253.5	45.6	257	21	V73360	HTM clone 2495790
3	71	12.8	205	19	W98718	H. pylori GHP0 808
4	68.5	12.3	413	20	Y07057	Renal cancer assoc
5	64.5	11.6	696	20	Y13359	Amino acid sequenc
6	64.5	11.6	696	21	Y70671	Human PRO266 prote
7	64	11.5	110	20	Y27041	Amino acid sequenc
8	63.5	11.4	755	20	W81365	Human prolactinome c
9	62	11.2	985	21	Y70073	Ross river virus e
10	61.5	11.1	720	17	R92227	Thermostable enzym
11	61	11.0	619	14	R42090	c-rel protein. Ho
12	61	11.0	721	18	W34454	Racillus subtilis

13	61	11.0	746	18	W34455	Racillus subtilis
14	60	10.8	410	20	W97090	A tairf enzyme of S
15	60	10.8	709	21	Y81802	Prostaglandin tran
16	60	10.8	709	21	Y77455	Human dendritic ce
17	59.5	10.7	260	21	Y81678	Streptococcus pneu
18	59.5	10.7	274	20	Y43931	Fission yeast prot
19	59.5	10.7	297	20	W95686	Schizosaccharomyce
20	59.5	10.7	398	20	Y34810	Chlamydia pneumoni
21	59.5	10.7	1001	13	R25068	hlir-R. Homo sapi
22	59.5	10.7	1001	15	R45774	Human leukemia in
23	59.5	10.7	1097	15	R74095	Human leukemia in
24	59.5	10.7	1097	15	R45058	Human LIF-R. Homo
25	59.5	10.7	1097	15	R45776	Human leukemia in
26	59.5	10.7	1097	16	R74097	Human leukemia in
27	59.5	10.7	1724	21	Y54373	CDNA sequence enco
28	59	10.6	360	20	Y06245	Human FAK-related
29	59	10.6	1052	17	R88576	Human focal adhesi
30	59	10.6	3567	14	R44431	erya region polype
31	58.5	10.5	153	20	Y29392	Sperm whale myoglo
32	58.5	10.5	201	19	W53964	Humicola insolens
33	58.5	10.5	201	19	W53975	Humicola insolens
34	58.5	10.5	201	19	W53976	Humicola insolens
35	58.5	10.5	201	19	W53977	Humicola insolens
36	58.5	10.5	201	19	W53978	Humicola insolens
37	58.5	10.5	234	19	W46618	Humicola insolens
38	58.5	10.5	286	19	W57420	Humicola insolens
39	58.5	10.5	286	20	Y37329	Protein involved I
40	58.5	10.5	305	12	R15271	Humicola insolens
41	58.5	10.5	305	13	R25525	Humicola insolens
42	58.5	10.5	305	13	R25464	Endoglucanase #1.
43	58.5	10.5	305	13	R25428	Cellulase containe
44	58.5	10.5	305	13	R27968	Endoglucanase enz
45	58.5	10.5	305	13	R28295	Sequence of ~ 43 k

ALIGNMENTS

RESULT	1	
ID	Y48540	standard; Protein; 212 AA.
XX	Y48540;	
AC	Y48540;	
XX	08-DEC-1999	(first entry)
DT	08-DEC-1999	
DE	Human breast tumour-associated protein 1.	
XX		
KW	Expressed sequence tag; EST; human; breast; cancer; gene therapy; treatment; tumour; cyostatic; medicament.	
OS	Homo sapiens.	
XX		
PN	DE19813839-A1.	
XX		
PD	23-SEP-1999.	
XX		
PF	20-MAR-1998;	98DE-1013839.
XX		
PR	20-MAR-1998;	98DE-1013839.
XX		
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.	
XX		
PI	Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;	
XX		
DR	WPI: 1999-528981/45.	
XX	N-PDB; Z33613.	
PT	Human nucleic acid sequences and protein products from tumor breast	
XX	tissue, useful for breast cancer therapy -	
PS	Claim 25; 143; 188pp; German.	

Query Match	45.6%	Score 253.5	DB 21	Length 257
Best Local Similarity	50.5%	Pred. No. 2.5e-23		
Matches 50	Conservative 20	Mismatches 26	Indels 3	Gaps 2
QY	4	LTGACVSVCSPGIPDERS-RDGIYARLAVDFDLPDQAMFDIEYRKPDPFFKFAKEITY 62		
Db	1	myvgagstpsgldptrpsgyslnl--qgydlpypeaifeipffhmkpfiftlakely 58		
QY	63	PGQFQPSLCIHKFIALSDKEGKILRRYQTQNIIDLEQVAGI 101		
Db	59	pgnykpnvthylfirlhbkgyllrllytqnlglervsgl 97		
RESULT 3				
W98718				
ID	W98718	standard; Protein; 205 AA.		
XX AC	W98718:			
XX DT	31-MAR-1999	(first entry)		
XX DE	H. pylori GHP0 808 protein.			
XX KW	GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;			
XX RM	peptic ulcer disease.			
XX OS	Helicobacter pylori.			
XX PN	W09843478-A1.			
XX PD	08-OCT-1998.			
XX PF	01-APR-1998;	98WO-US06371.		
XX PR	29-JUL-1997;	97US-0902615.		
XX PR	01-APR-1997;	97US-0833457.		
XX PR	24-JUN-1997;	97US-0861227.		
XX PA	(HUMA-) HUMAN GENOME SCI INC.			
XX PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.			
XX PI	Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;			
XX DR	WPI; 1998-542293/46.			
XX DR	N-PSDB; X14437.			
XX PT	New isolated Helicobacter polynucleotides - used to develop products			
XX PT	for the diagnosis, prevention and treatment of Helicobacter			
XX PT	infections and gastrointestinal diseases			
XX PS	Claim 8; Page 1517-1518; 2054pp; English.			

XX This sequence represents a Helicobacter pylori GPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.

XX Sequence 205 AA;

Query Match 12.8%; Score 71; DB 19; Length 205;

Best Local Similarity 31.9%; Pred. No. 0.57; Mismatches 22; Indels 12; Gaps 4;

DB 21 npqkvid--fyngrrrlf----evypnk-----ahkalelekyq-vnltqnvddlh 68
 QY 37 DPQAMDEYFRKDPPEFAKEIYPGQPSLCHKFIALSDKEGLLNRYQNTIDTLE 96
 DB 21 npqkvid--fyngrrrlf----evypnk-----ahkalelekyq-vnltqnvddlh 68
 QY 97 QVAGIORIL 105
 DB 69 eragssrll 77

RESULT 4

ID Y07057 standard; Protein; 413 AA.

AC Y07057;

DT 02-JUL-1999 (first entry)

DE Renal cancer associated antigen precursor sequence.

KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.

OS Homo sapiens.

PN W09904265-A2.

XX 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDWIG) INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;

XX WPI: 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure: Page 446-447; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX Sequence 413 AA;

Query Match 12.3%; Score 68.5; DB 20; Length 413;

Best Local Similarity 28.1%; Pred. No. 2.9; Mismatches 27; Indels 7; Gaps 3;

QY 2 IVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPQAMFIDYFRKDPPEFAKEI 61
 DB 167 illeelnaagrcrlpldes ntlhklviegqpppvaqe-ydvpfkdkedffnsgwl 224
 QY 62 YPGQFQPSL-----CHKFIALSDKEGLLNRYQNTI 92
 DB 225 ttgqilpyidgfrhgdksaadvelnlvrlaql 260

RESULT 5

ID Y13359 standard; Protein; 696 AA.

AC Y13359;

DT 25-JUN-1999 (first entry)

DE Amino acid sequence of protein PRO266.

KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.

XX Homo sapiens.

PN W09914328-A2.

XX 25-MAR-1999.

PF 16-SEP-1998; 98WO-US19330.

PR 25-NOV-1997; 97US-0066840.

PR 17-SEP-1997; 97US-0059113.

PR 17-SEP-1997; 97US-0059115.

PR 17-SEP-1997; 97US-0059117.

PR 17-SEP-1997; 97US-0059119.

PR 17-SEP-1997; 97US-0059121.

PR 17-SEP-1997; 97US-0059122.

PR 18-SEP-1997; 97US-0059184.

PR 18-SEP-1997; 97US-0059263.

PR 18-SEP-1997; 97US-0059266.

PR 15-OCT-1997; 97US-0062125.

PR 17-OCT-1997; 97US-0062287.

PR 21-OCT-1997; 97US-0063486.

PR 24-OCT-1997; 97US-0062814.

PR 24-OCT-1997; 97US-0062816.

PR 24-OCT-1997; 97US-0063045.

PR 24-OCT-1997; 97US-0063120.

PR 24-OCT-1997; 97US-0063121.

PR 24-OCT-1997; 97US-0063127.

PR 24-OCT-1997; 97US-0063128.


```

FT Modified-site 573..577
FT /note="Casein Kinase II phosphorylation site"
FT Modified-site 595..599
FT /note="N-glycosylation site"
FT Modified-site 598..604
FT /note="N-myristoylation site"
FT Modified-site 603..609
FT /note="N-myristoylation site"
FT Modified-site 608..612
FT /note="Casein Kinase II phosphorylation site"
FT Modified-site 646..650
FT /note="cAMP and cGMP-dependent protein kinase
FT phosphorylation site"
FT Modified-site 655..659
FT /note="N-glycosylation site"
FT Modified-site 657..661
FT /note="Casein Kinase II phosphorylation site"
FT Modified-site 666..670
FT /note="Casein Kinase II phosphorylation site"
FT Modified-site 693..697
FT /note="Casein Kinase II phosphorylation site"
XX MO200015797-A2.
XX 23-MAR-2000.
XX
XX 15-SEP-1999; 99WO-US21547.
XX
XX 17-SEP-1998; 98US-0100858.
XX 17-SEP-1998; 98WO-US19437.
XX
XX (GETH ) GENENTECH INC.
XX
XX Fong S, Goddard A, Gurney AL, Tunas D, Wood WI;
XX
XX WPI: 2000-271435/23.
XX N-PSDB; 252205.
XX
XX Composition for treatment and diagnosis of immune related diseases e.g.
XX Grave's disease comprises a PRO245, PRO217, PRO301, PRO266, PRO333,
XX PRO331 or PRO326 polypeptide or its agonists or antagonists (preferably
XX antibodies) -
XX
XX Example 1; Fig 10; 201pp; English.
XX
XX The present sequence is the human protein PRO266, encoded by UNQ23 cDNA,
XX designated as clone DNA37150. It is isolated from human foetal brain
XX tissue. PRO266 has significant homology to a SLIT protein, indicating
XX that it could be a leucine rich repeat protein.
XX It enhances or suppresses the infiltration of inflammatory cells into
XX tissues, proliferation of T-lymphocytes and modulates the immune
XX response. This sequence is useful for treatment of immune related
XX disorders, like SLE, rheumatoid/juvenile arthritis, spondyloarthritis,
XX systemic sclerosis (scleroderma), idiopathic inflammatory myopathies such
XX as dermatomyositis, Sjogren's syndrome, systemic vasculitis, sarcoidosis,
XX autoimmune haemolytic anaemia, thrombocytopenia, thyroiditis e.g. Grave's
XX disease, diabetes mellitus, immune-mediated renal disease e.g.
XX glomerulonephritis, demyelinating diseases such as multiple sclerosis and
XX Guillain-Barre syndrome, hepatobiliary diseases like hepatitis and
XX primary biliary cirrhosis, inflammatory and fibrotic lung diseases such
XX as inflammatory bowel disease (e.g. Crohn's disease), autoimmune or
XX immune-mediated skin diseases such as psoriasis, allergies like asthma,
XX immunological diseases of the lungs such as eosinophilic pneumonia and
XX transplantation associated diseases such as graft-versus-host-disease.
XX
XX Sequence 696 AA;

```

```

Query Match 11.6%; Score 64.5; DB 21; Length 696;
Best Local Similarity 25.3%; Pred. No. 18;
Matches 21; Conservative 13; Mismatches 24; Indels 25; Gaps 4;
OY 43 DIEFRKDPFFKFAKEIYPGQFO-----PSLCHKFTALS--DKEG 82

```

```

Db 131 dleylqgd-----fllldldpgafgdklevlllndllstlpavfgyypichldlrg 186
OY 83 KLEKNYQNIIDTLEQVAGIQRI L 105
Db 187 nrkltpye-evleqlpgiaeil 208

RESULT 7
Y27041
ID Y27041 standard; Protein; 110 AA.
XX
XX Y27041;
XX
XX 08-OCT-1999 (first entry)
XX
XX Amino acid sequence of rat cubilin CUB17 domain.
XX
XX Cubilin; epithelial glycoprotein receptor; vitamin B12; endocytosis;
XX toxicity; kidney; spleen; brain; liver; heart; thyroid; abortion; rat;
XX proteinuria; fetal malformation; fetal development; kidney damage; CUB.
XX
XX Rattus sp.
XX
XX WO9937757-A1.
XX
XX 29-JUL-1999.
XX
XX 21-JAN-1999; 99WO-US01259.
XX
XX 22-JAN-1998; 98US-0072197.
XX
XX (INRM ) INST NAT SANTE & RECH MEDICALE.
XX (TULA ) TULANE EDUCATIONAL FUND.
XX
XX Hammond TG, Verroust PJ;
XX
XX WPI: 1999-479045/40.
XX
XX New DNA encoding cubilin, used for treating toxicity, particularly
XX nephrotoxicity, and as marker of kidney damage
XX
XX Example 18; Fig 6B; 135pp; English.
XX
XX The invention relates to a rat cubilin protein. Cubilin is a ligand-
XX binding, epithelial glycoprotein receptor that facilitates uptake of
XX intrinsic factor/vitamin B12 complexes in intestines and kidney. It is
XX also involved in endocytosis and trafficking of light immunoglobulin
XX chains in renal proximal tubule cells. Host cells containing a vector
XX comprising the rat cubilin DNA sequence can be used for the recombinant
XX expression of the protein. Cubilin, or its fragments, are used to treat
XX or reduce toxicity, particularly in kidneys, spleen, brain, liver, heart
XX and thyroid. Cubilin mutations may also be implicated in idiopathic
XX proteinuria, fetal malformation, poor fetal development and spontaneous
XX abortions. Cubilin may also be used to raise specific antibodies, used
XX for its detection, or clones that express it, in standard immunoassays.
XX Fragments of cubilin DNA can also be used to detect cubilin mRNA in cell
XX and tissues, by hybridization. Abnormal levels of cubilin in the urine
XX are indicative of kidney damage. Sequences Y27036-51 represent CUB
XX domain sequences from rat cubilin and from homologous regions of other
XX proteins.
XX
XX Sequence 110 AA;

```

```

Query Match 11.5%; Score 64; DB 20; Length 110;
Best Local Similarity 21.7%; Pred. No. 1.8;
Matches 23; Conservative 19; Mismatches 46; Indels 18; Gaps 3;
OY 6 GAGVSVSCGIPDFRSRGIVARLAVDPDLPDQAMDIIEFRKDPFFKFAKEIYPGQ 65
Db 2 ggtvsgdsyve-----slgypLpYannvfqwfirglpglyltstfedlnlg 50

```

OY 66 FOPSLCHKFIAL--SDKEGKLIRNTQN-----IDTLEQVAGIQRI 104
 DB 51 ssgpctkdfveiwhtsgrvigrvycnsltpsvdtssnvasvkrv 96

RESULT 8

ID W81365 standard; Protein; 755 AA.
 AC W81365;
 XX
 XX 30-MAR-1999 (first entry)
 DE Human prothormone convertase 4.

XX Human prothormone convertase; testicular prothormone; testicular cell;
 KW fertility; spermatogenesis; embryo development; chromosome 19; 19p13.3.
 XX

OS Homo sapiens.

XX WO9850560-A1.

XX 12-NOV-1998.

XX 01-MAY-1998; 98WO-US08871.

XX 06-MAY-1997; 97US-0044015.

XX (ZYMO) ZYMOGENETICS INC.

PI Jaspers SR, Lok S;

DR WPI; 1999-059673/05.

DR N-PSDB; V84799.

XX New nucleic acid encoding human prothormone convertase 4 - useful
 PT for, e.g. identifying modulators and new testicular prothormones, and
 PT diagnosing chromosomal abnormalities

PS Claim 1; Page 70-75; 95pp; English.

XX Human prothormone convertase 4 can be used to raise antibodies; for in
 CC vitro identification of modulators and to identify, or determine
 CC function of, (new) testicular prothormones, and for processing of such
 CC prothormones (e.g. for stimulating proliferation or differentiation of
 CC testicular cells). mRNA for PC4 is detected only in testis, suggesting
 CC its involvement in fertility, spermatogenesis and early embryo
 CC development. The nucleic acid and its fragments (particularly probes and
 CC primers) are used to detect abnormalities in chromosome 19 (the gene for
 CC PC4 is located at 19p13.3).
 XX

SO Sequence 755 AA;

Query Match 11.4%; Score 63.5; DB 20; Length 755;
 Best Local Similarity 36.7%; Pred. No. 27;
 Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

OY 4 LTGAGVSYSCGIPDFRSRDGIYARLAVDFPDL---PDQAMDFYFRKDPFFKFAKE 60
 DB 147 lsgqglvvs--vld---dgl-----ekdhpdlwanydpdasydndydpdpqprytpsk 196

RESULT 9

ID Y70073 standard; Protein; 985 AA.

XX Y70073;

XX 05-JUN-2000 (first entry)

DE Ross river virus envelope glycoprotein from plasmid pRR64.
 XX

KW Ross river virus envelope glycoprotein; RRV; pRR64; viral glycoprotein;
 KW pseudotyped retrovirus; MMV; Moloney murine leukemia virus; antibody;
 KW transduction; screening agent; immunological agent;
 KW pharmacological agent.

OS Ross river virus.

XX WO200008131-A2.

XX 17-FEB-2000.

XX 04-AUG-1999; 99WO-US17702.

XX 04-AUG-1998; 98US-0095242.

XX 15-DEC-1998; 98US-0112405.

XX (PURD) PURDUE RES FOUND.

PI Sanders DA, Kuhn RJ, Jeffers SA, Sharkey CM, North CL;
 PI Fischbach MA;

DR WPI; 2000-224030/19.

DR N-PSDB; Z51037.

XX Cells that produce inventive pseudotyped retroviruses having a broad
 PT host range useful for introducing nucleotide sequences into target
 PT cells

XX Disclosure; Page 56-58; 65pp; English.

XX The present sequence is the Ross river virus (RRV) envelope
 CC glycoprotein. The coding region of this glycoprotein is amplified from
 CC plasmid pRR64. The amplified fragment contains the RRV E3-E2-6K-E1
 CC transmembrane glycoprotein coding region. This is used for generation of
 CC stable cell lines transiently producing RRV-MMV (Moloney murine
 CC leukemia virus) pseudotyped retrovirus with a broad host range. These
 CC cells having different viral glycoproteins in its lipid bilayer with a
 CC pseudotyped retrovirus is transduced with a desired ribonucleotide
 CC sequence. This is used to identify screening agents effective in
 CC blocking viral entry into a cell. These agents may be immunological
 CC agents like monoclonal or polyclonal antibodies. The pharmacological
 CC agents include proteins, peptides or various chemical agents. The
 CC pseudotyped retrovirus may be useful in methods of identifying cell
 CC surface receptors that allow viral entry.

SO Sequence 985 AA;

Query Match 11.2%; Score 62; DB 21; Length 985;
 Best Local Similarity 27.6%; Pred. No. 58;
 Matches 16; Conservative 10; Mismatches 28; Indels 4; Gaps 1;

OY 10 SYSCGIPDFRSRDGIYARLAVDFPDLPPQAMDFYFRKDPFFKFAKTIYP 63
 DB 327 nvtcrvlarapdvlygkvtlrlhpdhcltfsyrsigaephyeawkfserlhp 384

RESULT 10

ID R92227 standard; Protein; 720 AA.

XX R92227;

XX 14-MAY-1996 (first entry)

XX Thermostable enzyme.

XX Thermostable enzyme; trehalose; sweetener; saccharide.

XX Sulfolobus acidocaldarius strain ATCC 33909.

XX A09527132-A.

PD 01-FEB-1996.
 XX
 PF 21-JUL-1995; 95AU-0027132.
 XX
 PR 04-JUL-1995; 95UP-0189706.
 PR 21-JUL-1994; 94UP-0190183.
 XX
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX
 PI Kubota M, Maruta K, Sugimoto T;
 XX
 DR WPI; 1996-106285/12.
 DR N-PSDB; T15847.
 XX
 PT Recombinant thermostable enzyme from *Sulfolobus acidocaldarius*,
 PT forms non-reducing saccharide from reducing amylose saccharide,
 PT useful in trehalose prodn.
 XX
 PS Claim 2; Page 48-49; 70pp; English.
 XX
 CC A thermostable enzyme (R92227) is the product of a DNA sequence
 CC (T15847) isolated from *Sulfolobus acidocaldarius* ATCC 33909. The
 CC enzyme has a mol.wt. of 69-69 kDa, a pI of 5.4-6.4 and is
 CC substantially not inactivated when incubated at pH 7.0 and
 CC 85 deg.C for 60 min. The enzyme forms non-reducing saccharides
 CC having a trehalose end unit from reducing amylose saccharides.
 CC Recombinant thermostable enzyme is obtd. by expression of the
 CC encoding DNA in microbial transformants, esp. *Escherichia coli*.
 CC It is useful for commercial prodn. of trehalose, a sweetener.
 XX
 SQ Sequence 720 AA;

Query Match 11.1%; Score 61.5; DB 17; Length 720;
 Best Local Similarity 24.5%; Pred. No. 44;
 Matches 24; Conservative 18; Mismatches 47; Indels 9; Gaps 3;

OY 11 VSGIDPFRBDDIYARLANDFDLDPQAMFDIEYRKDPFRFRKAK---EIIYGOQ 66
 Db 576 msgidpdyggtelwlyld---pdrvrvpdkkhllekskstekmlesmhgdl 631
 OY 67 QPSLCHEKFIKLSDEKGLLRNTYQNTIDTLEQVAGIOR 103
 Db 632 kmlytkllslrqlaedflkgeykgldleaglcgflr 669

RESULT 11
 R42090
 ID R42090 standard; protein; 619 AA.
 XX
 AC R42090;
 XX
 DT 05-MAY-1994 (first entry)
 XX
 DE c-rel protein.
 XX
 KM NF-kappaB/Rel/Dorsal; NR; DNA binding region; human; hc-rel.
 KM
 OS Homo sapiens.
 OS
 PN MO9320219-A.
 PN
 PD 14-OCT-1993.
 PD
 PF 01-APR-1993; 93WO-US03027.
 PF
 PR 06-APR-1992; 92US-0862987.
 PR
 PA (USSH) US SEC DEPT HEALTH.
 PA
 PI Leonard WJ, Tolodano MB;
 PI
 DR WPI; 1993-336925/42.

XX
 PT Mutation of DNA binding region of NF-KB/rel/dorsal protein - to
 PT prevent or control binding to DNA for control of gene activation
 PT
 PS Disclosure; Page 36-37; 52pp; English.
 PS
 CC The sequence is that of the NF-kappaB c-rel or hc-rel protein, the DNA
 CC binding region of which may be mutated by replacing one or more amino
 CC acids. This allows inhibition, prevention or control of DNA binding and
 CC thus activation of genes controlled by the protein can be controlled.
 CC
 SQ Sequence 619 AA;

Query Match 11.0%; Score 61; DB 14; Length 619;
 Best Local Similarity 30.8%; Pred. No. 42;
 Matches 20; Conservative 5; Mismatches 22; Indels 18; Gaps 3;

OY 26 ARLAVDFDLDPQAMFDI---EYFRKDPFRFRKFA-----KELY---PQGFQ 67
 Db 336 agltvdiprprpql19sigegryfkkepnlfshdavrmpgvsgaesypspgpis 395
 OY 68 PSLCH 72
 Db 396 sg1sh 400

RESULT 12
 W34454
 ID W34454 standard; protein; 721 AA.
 XX
 AC W34454;
 XX
 DT 11-MAY-1998 (first entry)
 DT
 DE Racillus subtilis telchoic acid polymerase.
 DE
 XX

KM Telchoic acid polymerase; TAP enzyme; rod gene; assay.
 KM CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase;
 KW lipoteichoic acid; Gram-positive bacterium; infection;
 KW antimicrobial; antibiotic; screening.
 XX
 OS Bacillus subtilis.
 OS
 PN WO9742343-A2.
 PN
 PD 13-NOV-1997.
 PD
 PF 05-MAY-1997; 97WO-US07123.
 PF
 PR 07-MAY-1996; 96US-0016868.
 PR
 PA (PHAA) PHARMACIA & UPJOHN CO.
 PA
 PI Egan SE, Shinabarger DL, Swaney SM;
 PI
 DR WPI; 1997-558995/51.
 DR
 DR N-PSDB; T99933.
 DR
 XX

PT use of lipoteichoic acid as a substrate for new telchoic acid
 PT polymerase - to measure TAP activity, e.g. to monitor infection by
 PT gram-positive bacteria
 PT
 PS Claim 5; Page 35-37; 63pp; English.
 PS
 CC This sequence comprises telchoic acid polymerase (TAP) encoded by
 CC the rod gene (see T99933) of *Bacillus subtilis*. The use of
 CC lipoteichoic acid as a substrate for the reaction catalysed by TAP
 CC is claimed. Lipoteichoic acid, unlike telchoic acid, is
 CC commercially available and thus makes an excellent substrate for
 CC this important biological reaction that has previously had no
 CC substrate available. A new assay for measuring the activity of TAP
 CC comprises combining CDP-glycerol (which contains glycerol-3-phosphate),

CC the rodC gene (see T999936) of *Bacillus subtilis* cloned in pUC18.

XX

PT New Staphylococcus aureus CDP-glycerol:poly(glycerophosphate)
 PT glycerophospho transferase (tarf) enzyme polypeptide and
 PT polynucleotide - useful as diagnostic reagents and for prevention
 PT and treatment of Staphylococci infections
 PS Claim 12; Page 7; 26pp; English.

XX The present sequence represents a CDP-glycerol:poly(glycerophosphate)
 CC glycerophospho transferase (tarf) enzyme. The tarf enzyme is involved
 CC in the synthesis of wall teichoic acids. Tarf polypeptides and
 CC polynucleotides are useful for diagnosing diseases related to over or
 CC underexpression of tarf protein by identifying mutations in the tarf
 CC gene, or determining tarf polypeptide or mRNA expression levels due to
 CC an infection of an organism with the tarf gene. tarf polypeptides are
 CC also useful for their antagonists and agonists. tarf polypeptides can
 CC induce an immune response to immunise and prevent disease. Diseases
 CC diagnosed, prevented or treated include bacterial diseases (especially
 CC Staphylococcus aureus), including infections of the upper and lower
 CC respiratory tract (e.g. otitis media, thyroiditis), cardiac (e.g.
 CC infective endocarditis), gastrointestinal (e.g. secretory diarrhoea,
 CC splenic abscess), CNS (e.g. cerebral abscess), eye (e.g. conjunctivitis,
 CC keratitis), kidney and urinary tract (e.g. toxic shock syndrome), skin
 CC (e.g. impetigo, wound infection), and bone and joint (e.g. septic
 CC arthritis, osteomyelitis). Tarf polypeptides, polynucleotides and
 CC their (ant)agonists can prevent adhesion of bacteria to matrix proteins,
 CC and are useful for use on wounds and body implants to prevent bacterial
 CC infection.

SQ Sequence 410 AA;

Query Match 10.8%; Score 60; DB 20; Length 410;
 Best Local Similarity 23.7%; Pred. No. 32;
 Matches 18; Conservative 14; Mismatches 24; Indels 20; Gaps 2;

OY 25 YARLAVDFPDLPPQAM--FDIEYFRKDPPEFFAKETYPGOFPSLCHKFTALSDKEG 82
 : : | | | | :
 Db 317 yssvmtdfyglkrpqdyadldkgydelgfyfmdytkelp-----g 358
 OY 83 KLIRNTYQNTIDLEOV 98
 : : | | | | :
 Db 359 plvenhtalidaikqi 374

RESULT 15

ID Y81802 standard; Protein; 709 AA.

XX Y81802;

DT 07-JUN-2000 (first entry)

DE Prostaglandin transport homolog protein sequence.

KM Prostaglandin transporter homolog protein; PGTN; PGTN dysfunction;
 KM Prostaglandin transport activity; therapy.

XX Homo sapiens.

OS WO200009557-A1.

PN 24-FEB-2000.

PF 11-AUG-1999; 99WO-JP04352.

PR 12-AUG-1998; 98JP-0227723.

PA (KAZU-) KAZUSA DNA RES INST FOUND.
 PA (TAIS) TAISHO PHARM CO LTD.

PI Ohara O, Nagase T, Nomura N, Takayama K, Toyoda H, Yoshimoto M;
 DR WPI; 2000-205981/18.

DR N-PSDB; Z91909, Z91910.

XX New PGTN protein (prostaglandin transport homolog), used as a drug or
 PT in development of promoters or inhibitors of its gene, for treating
 PT PGTN dysfunction

PS Claim 1; Page 21-23; 33pp; Japanese.

XX This sequence is the human prostaglandin transport homolog (PGTN)
 CC protein of the invention. The protein has prostaglandin transport
 CC activity, and is useful in treating PGTN dysfunction. PGTN protein is
 CC used directly as a drug or in the development of agents such as promoters
 CC or inhibitors the pgtn gene, for treating diseases of PGTN dysfunction.

SQ Sequence 709 AA;

Query Match 10.8%; Score 60; DB 21; Length 709;
 Best Local Similarity 20.0%; Pred. No. 66;
 Matches 27; Conservative 21; Mismatches 45; Indels 42; Gaps 4;

OY 2 IVLTGAGVSVCIPDRSRDGIYARLAVDFPDLPPQAMFDIEYFRKDR----- 52
 : : | | :
 Db 233 vtmngpglafg-----smllrlyvdinqmpggsistl----kdrvwgawlgf 280
 OY 53 -----PPFKAKETYPGOFPSLCHKFTALSDKEGLRN-----YTQN 91
 : : | | :
 Db 281 llaagavalaalyfffpkempkerekelfgrrvavldsparkgkspksgpsgatk 340
 OY 92 IDTLEOVAGIQRITQ 106
 : : | | :
 Db 341 qdglvglaipnltylq 355

Search completed: February 16, 2001, 10:43:22
 Job time: 104 sec


```

Query Match 49.6%; Score 275.5; DB 2; Length 331;
Best Local Similarity 51.5%; Pred. No. 6e-22;
Matches 53; Conservative 20; Mismatches 27; Indels 3; Gaps 2;

.Oy 1 IIVLTGAGVSVSCGIPDERSRD-GIYARLAVDPDLDPDQAMFDIEYFRKDPPEFKPAK 59
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 VTFEFGAGISTGAGCIPDRSPDTGLYANLAK--LNLPEAAVDFIDPEKDPKPEYTLAE 78
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 60 ELYPGQFQPSLCHKRIALSDKEGKLLRNVTQNTIDPLEAGVQ 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79 ELYPGNFAPTKHHFTKLLODQGLSKRVYTONIDPLEKAGVE 121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
HST2 protein - yeast (Saccharomyces cerevisiae)
S59678
N:Alternate names: protein LPA2c; protein YPL015c
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C:Accession: S59678
R:Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friese, J.D.; Storms, R.K.; Vo, D.H.; We
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A:Reference number: S59677
A:Accession: S59678
A:Molecule type: DNA
A:Residues: 1-357 <HAL>
A:Cross-references: EMBL:U33335; NID:9965076; PIDN:AA868090.1; PID:9965078; MIPS:YPL015
C:Keywords: transmembrane protein
F:218-234/Domain: transmembrane #status predicted <TM>

Query Match 47.9%; Score 266.5; DB 2; Length 357;
Best Local Similarity 51.5%; Pred. No. 6e-21;
Matches 53; Conservative 18; Mismatches 29; Indels 3; Gaps 2;

Oy 1 IIVLTGAGVSVSCGIPDERS--RDGIYARLAVDFPDLDPDQAMFDIEYFRKDPPEFKPAK 59
      : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27 VTFEFGAGISTGAGCIPDRSPDTGLYHNLA--RLKLPPEAVDFIDPEKDPKPEYTLAK 84
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 60 ELYPGQFQPSLCHKRIALSDKEGKLLRNVTQNTIDPLEAGVQ 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 ELYPGNFRPSKFHYLLKLFQDKDVLKRVYTONIDPLEKAGVK 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
S59698
HST1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein O116l; protein YOL068c
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Jun-2000
C:Accession: S59698; S62108; S66761; S70376
R:Derbyshire, M.K.; Weinstock, K.G.; Strathern, J.N.
submitted to the EMBL Data Library, September 1995
A:Description: HST1, a new member of the SIR2 family of genes.
A:Reference number: S59698
A:Accession: S59698
A:Molecule type: DNA
A:Residues: 1-503 <DER>
A:Cross-references: EMBL:L47120; NID:9972892; PIDN:AA838430.1; PID:9972893
R:Brachmann, C.B.; Sherman, J.M.; Devine, S.E.; Cameron, E.E.; Pillus, L.; Boeke, J.D.
submitted to the EMBL Data Library, October 1995
A:Description: The SIR2 gene family, conserved from bacteria to humans, functions in silencing
A:Reference number: S62108
A:Accession: S62108
A:Molecule type: DNA
A:Residues: 1-503 <BRA>
A:Cross-references: EMBL:U39041; NID:91055019; PIDN:AA81033.1; PID:91055020
A:Alexandrakaki, D.; Katsoulou, C.; Tzermita, M.

```

[illegible]

transcription regulator SIR2 homolog - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999

C:Accession: T40929

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Aert, R.; Volckaert, G.

submitted to the EMBL Data Library, October 1999

A:Reference number: 221958

A:Accession: T40929

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-332 <LN>

A:Cross-references: EMBL:AL121807; PIDN:CAB58129.1; GSPDB:GN00068; SPDB:SPCC132.02

A:Experimental source: strain 972h-; cosmid c132

C:Genetics:

A:Gene: SPDB:SPCC132.02

A:Map position: 3

A:Introns: 27/1; 152/3

Query Match

43.8%; Score 243.5; DB 2; Length 332;

Best Local Similarity 49.0%; Pred. No. 1.6e-18;

Matches 50; Conservative 21; Mismatches 28; Indels 3; Gaps 2;

QY 1 IIVLTGAGVSVSCGIPDRSRD-GIYARLAVDPDLPQAMFDIEYFRKDPPEFFKAK 59

Db 30 ICWVWAGAGISIAAGIPDRSPETGTYNNL--QRENLPYAEAVFDLSYFRKDPPEFFELAH 87

QY 60 EIYFGQFQPSLCHKFIALSDKEGKLLRNNTONIDTLEOVAGI 101

Db 88 ELMEPEKYPPTYTHYFIRLLHDKRLLOKCYTONIDTLERLAGV 129

RESULT 7

Regulatory protein SIR2 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D2714; protein YDL042c; silent information regulator 2

C:Species: Saccharomyces cerevisiae

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000

C:Accession: S05891; S67575

R:Shore, D.; Squire, M.; Nasmyth, K.A.

EMBO J. 3, 2817-2823, 1984

A:Title: Characterization of two genes required for the position-effect control of yeast

A:Reference number: S05891; MUID:85126876

A:Accession: S05891

A:Molecule type: DNA

A:Residues: 1-562 <SHO>

A:Cross-references: EMBL:X01419; NID:g4469; PIDN:CAA25667.1; PID:g4470

R:Paulin, L.; Saren, A.M.; Laamanen, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67560

A:Accession: S67575

A:Molecule type: DNA

A:Residues: 1-562 <PAU>

A:Cross-references: EMBL:Z74090; NID:g1431026; PIDN:CAA96600.1; PID:g1431027; GSPDB:GN00

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:SIR2; MIPS:YDL042C

A:Cross-references: SGD:S0002200; MIPS:YDL042C

A:Map position: 4L

C:Superfamily: regulatory protein SIR2

C:Keywords: DNA binding; transcription regulation

Query Match

43.0%; Score 239; DB 1; Length 562;

Best Local Similarity 50.5%; Pred. No. 8.9e-18;

Matches 51; Conservative 12; Mismatches 36; Indels 2; Gaps 1;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPQAMFDIEYFRKDPPEFFKAK 60

Db 257 ILVLTGAGVSVSGIPDRSRDGIYARLAVDPDLPQAMFDIEYFRKDPPEFFKAK 314

QY 61 IYFGQFQPSLCHKFIALSDKEGKLLRNNTONIDTLEOVAGI 101

Db 315 VLPEKIVSPHSFIKMLQMKGKLLRNNTONIDNLESYAGI 355

RESULT 8

silent information regulator 2-related protein - Leishmania major

N:Alternate names: silent information regulator 2 protein homolog

C:Species: Leishmania major

C>Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 10-Sep-1997

C:Accession: JC4639; PC4152

R:Yahiaoui, B.; Taibi, A.; Ouasssi, A.

Gene 169, 115-118, 1996

A:Title: A Leishmania major protein with extensive homology to silent information reg

A:Reference number: JC4639; MUID:96186914

A:Accession: JC4639

A:Molecule type: mRNA

A:Residues: 1-381 <YAH>

A:Cross-references: GB:L40331; NID:g1203986; PID:g1203987

A:Accession: PC4152

A:Molecule type: protein

A:Residues: 58-76 <YAZ>

C:Genetics:

A:Gene: LmSIR2rp

C:Keywords: zinc finger

F:270-310/Region: serine-rich

F:152,155,176,179/Binding site: zinc (Cys) #status predicted

Query Match

42.9%; Score 238.5; DB 2; Length 381;

Best Local Similarity 51.0%; Pred. No. 6.5e-18;

Matches 53; Conservative 14; Mismatches 32; Indels 5; Gaps 3;

QY 1 IIVLTGAGVSVSCGIPDRSRD-GIYARLAVDPDLPQAMFDIEYFRKDPPEFFKAK 59

Db 34 ILVLTGAGVSVAGIPDRSSDTGTIYAKLGK--YVNDPTDAFSTLLREKPEIFYSIAR 91

QY 60 E-IYFGQFQPSLCHKFIALSDKEGKLLRNNTONIDTLEOVAGI 101

Db 92 ELNMPGHQPTAVHFTIRLLQDEGRLLRCTONIDGLEKAGV 135

RESULT 9

T39571

probable regulatory protein sir2-like - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39571

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.

submitted to the EMBL Data Library, March 1999

A:Reference number: Z21864

A:Accession: T39571

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-471 <WOO>

A:Cross-references: EMBL:AL035637; PIDN:CAB38511.2; GSPDB:GN00067; SPDB:SPBC16D10.07C

A:Experimental source: strain 972h-; cosmid c16D10

C:Genetics:

A:Gene: SPDB:SPBC16D10.07C

A:Map position: 2

A:Introns: 53/1; 79/1; 106/3; 117/2; 263/1

Query Match

40.6%; Score 226; DB 2; Length 471;

Best Local Similarity 46.4%; Pred. No. 1.8e-16;

Matches 51; Conservative 19; Mismatches 32; Indels 8; Gaps 4;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPQAMFDIEYFRKDPPEFFKAK 60

Db 155 VVLTGAGVSVSGIPDRSDNGFYARLARH--GLSEPEMFDIHTRENDIEFYTFARD 212

QY 61 IYFGQFQPSLCHKFIALSDKEGKLLRNNTONIDTLEOVAGI--QRILO 106

Db 213 LLPEINHSFSS--HAFIRLEKKNKLSLTFTONIDNLEKKTGSLDNKTIQ 260

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
 A:Reference number: A69250; MWID:98049343
 A:Accession: H69263
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-253 <RLE>
 A:Cross-references: GB:AE001098; GB:AE000782; NID:g2689421; PIDN:AAB91115.1; PID:g265053
 C:Superfamily: conserved hypothetical protein b1120

Query Match 31.4%; Score 174.5; DB 2; Length 253;
 Best Local Similarity 36.8%; Pred. No. 2,9e-11;
 Matches 39; Conservative 22; Mismatches 36; Indels 9; Gaps 2;

QY 2 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPPQAMFDIEYFRKDPFFKFAKEI 61
 Db 19 VVFTGAGISAESGIPYFRGEGDLMWKY-----DPEEVAISGCFRNPRAEFSEMEM 70

QY 62 YPGQF--PSLCHKFTALSDEKGLRNVTQNTIDTLEQVAGIQRIIQ 106
 Db 71 KDKLFAPNPANHAIALAELERMGIKAVITQNTIDMLHGRAGSRVLE 116

RESULT 14
 T50106
 hst4p [imported] - fission yeast (Schizosaccharomyces pombe) .
 C:Species: Schizosaccharomyces pombe
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
 C:Accession: T50106
 R:Seeger, K.; Harris, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, January 2000
 A:Reference number: 225038
 A:Accession: T50106
 A:Status: preliminary; translated from GB/EMBL/DDB
 A:Molecule type: DNA
 A:Residues: 1-415 <SEE>
 A:Cross-references: EMBL:AL136499; PIDN:CAB6167.1; GSPDB:GNO00066; SPDB:SPAC1783.04C
 A:Experimental source: strain 972h(-); cosmid c1783
 C:Genetics:
 A:Gene: SPDB:SPAC1783.04C
 A:Map position: 1
 A:introns: 65/1

Query Match 29.2%; Score 162.5; DB 2; Length 415;
 Best Local Similarity 36.4%; Pred. No. 9.9e-10;
 Matches 39; Conservative 24; Mismatches 39; Indels 5; Gaps 3;

QY 1 IIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPPQAMFDIEYFR--KDPFRFFKFA 58
 Db 60 IIVVTTGAGISCDAGIPDFRSSEGLFSLRAEYKINCSEKELFDGSVYRDLKSVNIFHAMI 119

QY 59 KEIY--PGQFQPSLCHKFTALSDEKGLRNVTQNTIDTLE-QVAGIQ 102
 Db 120 RKLHMLSNRNPDPDFHFLSQLAQESKLRLYQNTIDLEFLERLEGLQ 166

RESULT 15
 D72562
 hypothetical protein APE1782 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: D72562
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatake, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A:Reference number: A72450; MWID:99310339
 A:Accession: D72562
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-247 <RAW>
 A:Cross-references: DDB:AP000062; NID:g5105244; PIDN:BAA80785.1; PID:g5105472

A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1782
 C:Superfamily: conserved hypothetical protein b1120

Query Match 27.5%; Score 153; DB 2; Length 247;
 Best Local Similarity 32.4%; Pred. No. 5.7e-09;
 Matches 36; Conservative 20; Mismatches 37; Indels 18; Gaps 3;

QY 2 IVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPPQAMFDIEYFRKDPFFKFAKEI 61
 Db 20 VAFRAGISAESGIPYFRGKGLMSRF-----DPRDLATPEAFNRDPR----LVWEM 67

QY 62 YPGQFQ-----PSLCHKFTALSDEKGLRNVTQNTIDTLEQVAGIQRIIQ 106
 Db 68 YSMRIENVLAAPKPAHRLARLEDSGLKAVITQNTIDGLHRRAGSRVLE 118

Search completed: February 16, 2001, 10:45:30
 Job time: 231 sec

DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 SN SIR2 OR MAR1 OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85126876; PubMed-6098447;
 RA Shore D., Squire M., Nasmyth K.A.;
 RT "Characterization of two genes required for the position-effect
 RT control of yeast mating-type genes.";
 RL EMO J. 3:2817-2823(1984).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Paulin L., Saren A.M., Laamanen P.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: X01419; CA25667.1; -
 DR EMBL: Z71781; CA96447.1; -
 DR EMBL: Z74090; CA98600.1; -
 DR PIR: S05891; RCBY52.
 DR SGD: S0002200; SIR2.
 DR Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 254 498 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 372 399 C4-TYPE (POTENTIAL).
 FT SEQUENCE 562 AA; 63261 MW; 52E6937533654586 CRC64;
 SQ
 Query Match 43.0%; Score 239; DB 1; Length 562;
 Best Local Similarity 50.5%; Pred. No. 5e-19;
 Matches 51; Conservative 12; Mismatches 36; Indels 2; Gaps 1;
 QY 1 IIVLTGAGVSVSGGIPDRSRDGIYARLAVDPDLPDQAMPDIEYFRKDPPEFFAKE 60
 DB 257 ILVLTGAGVSTSLGIPDRSSEGFYSK-IKHLGLDDPDQVFNINIMHDPFSVFNINAM 314
 QY 61 IYPGQFQPSLCHKEFIALSDKEGKLLRNVTQNTIDPLEOVAGI 101
 DB 315 VLPEKITYPLHSFTIKMLQMKGLLRNTYQNTIDNLESYAGI 355
 RESULT 5
 SIR2_LEIMA
 ID SIR2_LEIMA STANDARD; PRT; 381 AA.
 AC Q25337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RP).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96186914; PubMed-8635734;
 RA Yahlaoui B., Talbl A., Ouasssi A.;

RT "A Leishmania major protein with extensive homology to silent
 RT information regulator 2 of Saccharomyces cerevisiae.";
 RL Gene 169,115-118(1996).
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: I40331; AAB06804.1; -
 DR Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
 FT ZN-FING 148 151 POLY-ALA.
 FT ZN-FING 152 179 C4-TYPE (POTENTIAL).
 FT DOMAIN 271 278 POLY-SER.
 FT DOMAIN 305 310 POLY-SER.
 FT SEQUENCE 381 AA; 41958 MW; 5A311630A5D2365 CRC64;
 SQ
 Query Match 42.9%; Score 238.5; DB 1; Length 381;
 Best Local Similarity 51.0%; Pred. No. 3.7e-19;
 Matches 53; Conservative 14; Mismatches 32; Indels 5; Gaps 3;
 QY 1 IIVLTGAGVSVSGGIPDRSRDGIYARLAVDPDLPDQAMPDIEYFRKDPPEFFAKE 59
 DB 34 ILVLTGAGVSAAGIPDRSSDTGIYAKLCK-YNLDDPDAFSLTLREKPEIFYSIAR 91
 QY 60 E-IYPGQFQPSLCHKEFIALSDKEGKLLRNVTQNTIDPLEOVAGI 101
 DB 92 ELNLMPGHQPTAVHPIRLLDQEGRLRCCCTQNTIDLEKAAVY 135
 RESULT 6
 SIR2_CANAL
 ID SIR2_CANAL STANDARD; PRT; 515 AA.
 AC Q59923;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE REGULATORY PROTEIN SIR2.
 GN SIR2.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC anamorphic Saccharomycetales; Candida.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martin J., Johnson A.D.;
 RT "SIR2 gene from Candida albicans.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@lsb-sib.ch.

CC -----
 DR EMBL: AF045774; AAC09304.1; -
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 FT Nucleic protein.
 FT DOMAIN 234 453 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 352 379 C4-TYPE (POTENTIAL).
 SO SEQUENCE 515 AA; 57732 MW; 284DB259FEA9251F CRC64;

Query Match 42.1%; Score 234; DB 1; Length 515;
 Best Local Similarity 49.0%; Pred. No. 1.6e-18;
 Matches 50; Conservative 19; Mismatches 31; Indels 2; Gaps 1;

OY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDPDPOAMFDIEFRKDPPEFFAKE 60
 DB 237 IMVYTGAGISTSLGIPDRSRFGIXNLSK--LNSDPQKVPDLOTFRREGLEFTIAHL 294
 OY 61 IYPCGFQPSLCHKFTALSDKREGKLLRNTQNTIDTLEQVAGI 102
 DB 295 VLPDGRFSLHAFLLQDKHKLRLRYNTQNTIDLEQVAGI 336

RESULT 7
 HST3_YEAST STANDARD; PRT; 447 AA.
 AC P53687;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST3 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
 GN HST3 OR YOR025W OR OR26.15.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomycetes.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPH1;
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RA de Haan M., Maarse A.C., Grivell L.A.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
 PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
 AND GENOMIC MAINTENANCE.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: U39062; AAA81034.1; -
 DR EMBL: X87331; CAA60741.1; -
 DR EMBL: Z74933; CAA99215.1; -
 DR SGD: S0005551; HST3.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nucleic protein.
 FT DOMAIN 50 314 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 195 223 C4-TYPE (POTENTIAL).
 SO SEQUENCE 447 AA; 50524 MW; 59CDSFBD12B7005 CRC64;

Query Match 35.0%; Score 194.5; DB 1; Length 447;
 Best Local Similarity 44.4%; Pred. No. 3.5e-14;
 Matches 48; Conservative 14; Mismatches 39; Indels 7; Gaps 3;

OY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDPDPOAMFDIEFRKDPPEFFAKE 55
 DB 55 IACLTGAGISNACGIPDRSRDGLDYLKDCSQYWSIKSGREMDISLFRDDEKISIFA 114
 OY 56 KEAFKIYPC--QFQPSLCHKFTALSDKREGKLLRNTQNTIDTLEQVAGI 101
 DB 115 KFERLYSNVOLAKPTKTHKFTALHKLDRKKLLRNTQNTIDGLESTIGL 162

RESULT 8
 HST4_YEAST STANDARD; PRT; 370 AA.
 AC P53688;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 4).
 GN HST4 OR YDR191W OR YD9346.03.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomycetes.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP GENE NAME.
 RC STRAIN-GRY 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).

RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995)
 CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST3 TO TELOMERIC SILENCING,
 PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
 AND GENOMIC MAINTENANCE.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: Z48784; CAA88705.1; -
 DR SGD: S0002599; HST4.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nucleic protein.
 FT DOMAIN 90 342 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 221 254 C4-TYPE (POTENTIAL).
 SO SEQUENCE 370 AA; 41765 MW; 86BBO238BFA914F1 CRC64;

Query Match 25.5%; Score 142; DB 1; Length 370;
 Best Local Similarity 35.6%; Pred. No. 1.9e-08;
 Matches 36; Conservative 22; Mismatches 33; Indels 10; Gaps 3;

QY 1 IIVLTGAGVSVSCGIPDFSRSDGIYARL-AVDEPDLPPQAMFIDE---YFRKDPFRFF 55
 DB 95 MIVVSGAGISVANGIPDFSRSEGIPTVNGSGSKDLFDYRVRVYGDSEMSLKFQNLWMSLF 154
 QY 56 KRAKEIYPCGQPSLCHKFIALSDKEGKLLRNTQNTIDLE 96
 DB 155 RLSKNCOPTKF-----HEMLNEFARDGRLLRNTQNTIDGID 190

RESULT 9

Y104_MYCTU
 ID Y104_MYCTU STANDARD: PRT; 504 AA.
 AC 010898;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 53.5 KDA PROTEIN RV0104.
 GN RV0104 OR MTCY251.23.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brochier R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogan A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sultson J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z74410; CAA98917.1; -
 CC HSSP: P00514; IBPK.
 DR TUBERCULIST; RV0104; -
 DR INTERPRO: IPR000595; -
 DR PFAM: PF00027; CNMP_binding; 1.
 DR PROSITE: PS00042; CNMP_BINDING_3; UNKNOWN_1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 146 166 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 SQ SEQUENCE 504 AA; 53450 MW; 3FCF8A42EE75009F CRC64;

Query Match 12.1%; Score 67.5; DB 1; Length 504;
 Best Local Similarity 26.5%; Pred. No. 5.4;
 Matches 27; Conservative 11; Mismatches 55; Indels 9; Gaps 2;

QY 2 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPPQAMFIDEYFRKDPFRFFKAKEI 61
 DB 53 VAVIGACYSHPGVDMRADGIYD---DCSDAYAPHESEFDQYTRHVERFEAESWARL 109
 QY 62 YPGQFQPSLCHKFIALSDKEGKLLRNTQNTIDLEQVAGIOR 103
 DB 110 TAGRTAR-----VLLDDGSLAVAGAMLADASADYIGIQ 145

RESULT 10

THRC_HELPY
 ID THRC_HELPY STANDARD: PRT; 486 AA.
 AC 024924;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THREONINE SYNTHASE (EC 4.2.99.2).
 GN THRC OR HP0098.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695 / ATCC 700392;
 RX MEDLINE-97394467; PubMed-9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Meldman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori."
 RL Nature 388:539-547(1997).
 CC -1- CATALYTIC ACTIVITY: O-PHOSPHO-L-HOMOSERINE + H(2)O = L-THREONINE +
 CC ORTHOPHOSPHATE.
 CC -1- PATHWAY: THREONINE BIOSYNTHESIS.

CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE000531; AAD07166.1; -
 CC TIGR: HP0098; -
 DR PFAM: PF00291; S_T_dehydratase; 1.
 DR PROSITE: PS00165; DEHYDRATASE_SRR_THR; 1.
 KW Threonine biosynthesis; Lyase; Pyridoxal phosphate.
 FT BINDING 109 109 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 SQ SEQUENCE 486 AA; 54706 MW; CD9637799B508988 CRC64;

Query Match 12.0%; Score 66.5; DB 1; Length 486;
 Best Local Similarity 28.2%; Pred. No. 6.7;
 Matches 29; Conservative 14; Mismatches 35; Indels 25; Gaps 5;

QY 24 IYALAVDPDLPPQAMFIDEYF-RKDPFRFFK-----FAKEIYPCGQPSLCHKFIA 76
 DB 56 VFERLGLFIRKNLALSKRRENDKPNRPAIFALMERLFVQELIYG---PSLAFDMA 112
 QY 77 LS-----DKEGKLLRNTQNTID---TLEQVAGIORI 104
 DB 113 LQPLASLFSNLAVGKNKRYLMLVTSDDTGPATLESILAGMNV 155

RESULT 11
 YE20_METJA
 ID YE20_METJA STANDARD: PRT; 1102 AA.
 AC 058815;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MJ1420.
 GN MJ1420.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

```

CC Methanococcus .
RN [1]
RP SEQUENCE FROM N.A. / DSM 2661 / ATCC 43067;
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Butt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Georgannas N.S.M., Weidman J.F., Furumann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RT Science 273:1058-1073(1996).
RL
CC -1- PWM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: IN THE C-TERMINAL, NO GLUCOSAMINE--FRUCTOSE-6-
CC PHOSPHATE AMINOTRANSFERASES (GFAT).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67582; AAB99430.1; -.
CC DR HSSP: P17169; IMOR.
CC DR MEROPS: C44.971; -.
CC DR TIGR: MJ1420; -.
CC DR INTERPRO: IPRO00583; -.
CC DR INTERPRO: IPRO01347; -.
CC DR INTERPRO: IPRO01387; -.
CC DR INTERPRO: IPRO02203; -.
CC DR PFAM: PF00310; GATase_2; 2.
CC DR PFAM: PF01381; HTH_3; 1.
CC DR PFAM: PF01380; SIS; 2.
CC DR PROSITE: PS00443; GATASE_TYPE_II; 1.
CC DR PROSITE: PS00881; PROTEIN_SPLICING; FALSE_NEG.
CC KW Hypothetical protein: protein splicing.
CC FT CHAIN 1 74
CC FT CHAIN 75 573
CC FT CHAIN 574 1102
CC FT CHAIN 1102 125907 MW; 9CIEF88BD90CBA1 CRC64;
CC SO SEQUENCE 1102 AA; 125907 MW; 9CIEF88BD90CBA1 CRC64;
CC -----
Query Match 11.9%; Score 66; DB 1; Length 1102;
Best Local Similarity 27.5%; Pred. No. 20;
Matches 22; Conservative 11; Mismatches 31; Indels 16; Gaps 3;
QY 21 RDGIYARLAVDPDLDPQAMFDIEYFKKDPREFKAKETIYPGQFQSLCHKETALS DK 80
Db 259 RDNLHYTPNVPKPKPTPELMQIIIGIYID-----GHF-PS--NRMLRLKDE 302
QY 81 ESKLRLANTQNDLTLEQVAG 100
Db 303 RKEVLEEYNQLEKTVENLEGG 322
RESULT 12
PMIP_YEAST ID PMIP_YEAST STANDARD; PRT; 772 AA.
AC P33599; P31980;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.55) (MIP).

```

```

OS OCT1 ORMPL OR YKLJ34C.
SN Saccharomyces cerevisiae (Baker's yeast)..
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94309680; PubMed=8035833;
RA Isaya G., Miklos D., Rollins R.A.;
RT "MPL, a new yeast gene homologous to the rat mitochondrial
RT intermediate peptidase gene, is required for oxidative metabolism in
RT Saccharomyces cerevisiae."
RL Mol. Cell. Biol. 14:5603-5616(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Rad M.R., Xu G., Kirchath L., Filtz C., Keuchel H., Hollenberg C.P.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO
CC THEIR MATURE SIZE. CLEAVES THE NUCLEAR-ENCODED PRECURSORS FOR
CC CYTOCHROME OXIDASE SUBUNIT IV (COXI) AND THE IRON-SULFUR PROTEIN
CC (FE-S) OF THE BC1 COMPLEX.
CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND
CC STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE
CC MITOCHONDRION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U10243; AAA21278.1; -.
CC DR EMBL: Z28134; CAAB1975.1; -.
CC DR PIR: S37963; S37963.
CC DR MEROPS: M03.006; -.
CC DR SGD: S0001617; OCT1.
CC DR INTERPRO: IPR000130; -.
CC DR INTERPRO: IPR001567; -.
CC PFAM: PF01432; Peptidase_M3; 1.
CC DR PROSITE: PS00142; ZINC_PROTEASE; 1.
CC KW TRANSIT: Metalloprotease; zinc; Zinc; Translt peptide; Mitochondrion.
CC FT CHAIN 1 37 MITOCHONDRION (POTENTIAL).
CC FT METAL 38 772 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 558 559 559 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 562 562 562 BY SIMILARITY.
CC FT METAL 565 565 565 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT CONFLICT 343 350 MAKNPBDV -> WQDRRC (IN REF. 2).
CC FT CONFLICT 694 694 Y -> S (IN REF. 2).
CC FT CONFLICT 696 702 ATYSYL -> QLTATY (IN REF. 2).
CC SEQUENCE 772 AA; 88182 MW; 1C19A0655FAAE7CA CRC64;

Query Match 11.8%; Score 65.5; DB 1; Length 772;
Best Local Similarity 29.2%; Pred. No. 15;
Matches 21; Conservative 10; Mismatches 28; Indels 13; Gaps 3;

QY 27 RLAVDFPLLDPQAMEDIYFRKDRPFPPFAKETIYPGQ-FQPSLCIKHFIALSDKEGKLL 85
Db 580 RCATFVEFLPS-----ILMEHFAKDRIRLTIKGKHGTGETIQADMLQIFPKMST----- 628
QY 86 RNYYQNIDTLQG 97
Db 629 -NFLQNCETYSQ 639

RESULT 13
M3KA_HUMAN STANDARD: PRT; 954 AA.
ID M3KA_HUMAN
```

AC 002779; Q12761; Q14871;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MITOGEN-ACTIVATED PROTEIN KINASE KINASE 10 (EC 2.7.1.1-)
 DE (MIXED LINEAGE KINASE 2) (PROTEIN KINASE MST).
 GN MAPK10 OR MLK2 OR MST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-96128179; PubMed-8536694;
 RA Dorow D.S., Devereux L., Tu G.F., Price G., Nicholl J.K.,
 RA Sutherland G.R., Simpson R.J.,
 RT "Complete nucleotide sequence, expression, and chromosomal
 RT localisation of human mixed-lineage kinase 2.";
 RL Eur. J. Biochem. 234:492-500(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-95249256; PubMed-7731697;
 RA Kato M., Hirai M., Sugimura T., Terada M.,
 RT "Cloning and characterization of MST, a novel (putative)
 RT serine/threonine kinase with SH3 domain.";
 RL Oncogene 10:1447-1451(1995).
 RN [3]
 RP SEQUENCE OF 244-480 FROM N.A.
 RC TISSUE-COLON EPITHELIAL;
 RX MEDLINE-9338756; PubMed-8477742;
 RA Dorow D.S., Devereux L., Dietzsch E., de Kreijter T.,
 RT "Identification of a new family of human epithelial protein kinases
 RT containing two leucine/isoleucine-zipper domains.";
 RL Eur. J. Biochem. 213:701-710(1993).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND SKELETAL MUSCLE.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE KINASE SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X90846; CAA62351.1; -
 DR EMBL: 248615; CAA88531.1; -
 DR PIR: S32468; S32468.
 DR HSSP: P00523; 2PTK.
 DR MIM: 600137; -
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR001245; -
 DR INTERPRO: IPR001452; -
 DR INTERPRO: IPR002965; -
 DR PFAM: PF00018; SH3; 1.
 DR PFAM: PF00069; pkinase; 1.
 DR PRINTS: PRO0109; TYRKINASE.
 DR PRINTS: PRO0452; SH3DOMAIN.
 DR PRINTS: PRO1217; PRICHEXTENSIN.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; SH3 domain.
 FT DOMAIN 2 5
 FT DOMAIN 16 81
 FT DOMAIN 98 360
 FT NP_BIND 104 112
 FT BINDING 125 125
 ATP (BY SIMILARITY).
 ATP (BY SIMILARITY).
 ATP (BY SIMILARITY).

FT ACT SITE 222 222 BY SIMILARITY.
 FT DOMAIN 384 405 LEUCINE-ZIPPER (BY SIMILARITY).
 FT DOMAIN 419 440 LEUCINE-ZIPPER (BY SIMILARITY).
 FT DOMAIN 449 463 ARG/LYS-RICH (BASIC).
 FT CONFLICT 462 464 SRL -> AV (IN REF. 2).
 FT CONFLICT 471 471 G -> S (IN REF. 2).
 FT CONFLICT 807 807 G -> R (IN REF. 2).
 FT CONFLICT 818 818 V -> A (IN REF. 2).
 FT CONFLICT 465 480 LKREGSHSLPSGF -> AQAGRRQHPALWL (IN
 FT REF. 3).
 SQ SEQUENCE 954 AA; 103623 MW; 538F4AA559B0ABA CRC64;
 Query Match 11.8%; Score 65.5; DB 1; Length 954;
 Best Local Similarity 38.3%; Pred. No. 19;
 Matches 18; Conservative 4; Mismatches 16; Indels 9; Gaps 2;
 OY 30 VDFPDLPDPQAMPDIEYFRKDPPE-----FKFAKITYGQPPQPSL 70
 DB 853 LDFPRLPDPQALFPA---RRRPFEGFRPTTLTFAPRRAPASRRL 896
 RESULT 14
 DD8 CAEEL STANDARD; PRT; 1200 AA.
 ID DD8 CAEEL
 AC 009530;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE EED8.5.
 GN EED8.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-BRISTOL N2;
 RA Chissee S.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD ACT LATE IN THE SPLICING OF PRE-MRNA AND MEDIANE
 CC THE RELEASE OF THE SPLICED RNA FROM SPLICEOSOMES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES. BELONGS TO THE
 CC "DEAD" SUBFAMILY. DD8/PRP22 ORTHOLOG.
 CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U23484; AAC46765.1; -
 DR HSSP: P05035; ISRO.
 DR WORMPEP: EED8.5; CE01889.
 DR INTERPRO: IPR002464; -
 DR INTERPRO: IPR003029; -
 DR PFAM: PF00575; SI; 1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE; 1.
 KW Hypothetical protein; mRNA processing; mRNA splicing; Helicase;
 KW ATP-binding; Nuclear protein.
 FT DOMAIN 176 185 POLY-ARG.
 FT DOMAIN 234 304 SI MOTIF.
 FT NP_BIND 563 570 ATP (POTENTIAL).
 FT SITE 660 663 DEAD BOX.
 SQ SEQUENCE 1200 AA; 135759 MW; 023C46E94D7F44F CRC64;
 Query Match 11.8%; Score 65.5; DB 1; Length 1200;
 Best Local Similarity 24.6%; Pred. No. 24;
 Matches 28; Conservative 21; Mismatches 48; Indels 17; Gaps 4;

OY 4 LTAGVSVSCGIPDFRSRDGIYARLAVDPDLPDQAMFIDIEYFRKDP---RPFFKFAKE 60
DB 376 MGGAGVLTATMDPDEDEKGVLR----NYDDESGEDI-EILVDEDEPFLGYKGGAE 430
OY 61 IYPGPF-----QPSLCHKFIALSDKEGKLLRNNTONITFLQVACIOIIL 105
DB 431 IEPVRVKNPDGSLAALMGGALSKERETKIQAQREHMDTOKGFSSNARIL 484

RESULT 15

HXA5_RAT STANDARD: PRT: 233 AA.

AC P52949;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HOMEOBOX PROTEIN HOX-A5 (HOX-1.3) (FRAGMENT).
GN HOXA5 OR HOXA-5 OR HOX-1.3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTA;
RX MEDLINE=94347374; PubMed=7915120;
RA Gorski D.H., Lepage D.F., Walsh K.;
RT "Cloning and sequence analysis of homeobox transcription factor cDNAs
with an inosine-containing probe."
RL Biotechniques 16:856-858(1994).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC ALSO BINDS TO ITS OWN PROMOTER. BINDS SPECIFICALLY TO THE MOTIF:
5'-CYYNATTA(TG)G-3'
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L03556; AAA67844.1; -
DR HSP: P02833; ISAN.
DR INTERPRO: IPR001356; -
DR INTERPRO: IPR001827; -
DR PFAM: PF00046; homeobox; 1.
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS00032; ANTENNAPEDIA; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT NON_TER 1 1
FT DOMAIN 155 160 ANTP-TYPE HEXAPEPTIDE.
FT DNA_BIND 174 233 HOMEOBOX.
FT NON_TER 233 233
SQ SEQUENCE 233 AA; 25387 MW; 0937608EEDAF368C CRC64;

Query Match 11.7%; Score 65; DB 1; Length 233;

Best Local Similarity 26.7%; Pred. NO. 4.2; Mismatches 49; Indels 6; Gaps 2;

OY 7 AGVSVSCGIPDFRSRDGIYARLAVDPDLPDQAMFIDIEYFRKDPPEFFKFAKEIYPCQF 66
DB 105 SGCANCGSTHSSREGVGTASAAE---EDAPASSEQAGAGSERPAPPAQPIYPMNR 160
OY 67 QPSLCHKFIALSDKEGKLLRNNTONITFL 96

DB 161 KLISHDNI--GGPEGRARCTYTRYQTL 188

Search completed: February 16, 2001, 10:52:36
Job time: 339 sec

REFERENCE/DOCKET NUMBER: 97-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 755 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-071-101-2

Query Match 11.4%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

QY 4 LTGAGVSCGIDPFRSDGIYARLAVDPDL---PPQAMFDEIYFRKDPREFKFAKE 60
Db 147 LSGGGIVS--VLD---DGI---EKDHPDLNMYDPLASYDNDYDPDPQPRYSKE 196

RESULT 5
US-09-369-618-2
Sequence 2, Application US/09369618
Patent No. 6100041
GENERAL INFORMATION:
APPLICANT: Lok, S1
APPLICANT: Jaspers, Stephen R.
TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
FILE REFERENCE: 97-05D2
CURRENT APPLICATION NUMBER: US/09/369,618
CURRENT FILING DATE: 1999-08-06
EARLIER APPLICATION NUMBER: US 09/071,101
EARLIER FILING DATE: 1998-05-01
EARLIER APPLICATION NUMBER: US 60/044,015
EARLIER FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 755
TYPE: PRT
ORGANISM: Homo sapiens
US-09-369-618-2

Query Match 11.4%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;
QY 4 LTGAGVSCGIDPFRSDGIYARLAVDPDL---PPQAMFDEIYFRKDPREFKFAKE 60
Db 147 LSGGGIVS--VLD---DGI---EKDHPDLNMYDPLASYDNDYDPDPQPRYSKE 196

RESULT 6
US-09-369-617-2
Sequence 2, Application US/09369617
Patent No. 6127162
GENERAL INFORMATION:
APPLICANT: Lok, S1
APPLICANT: Jaspers, Stephen R.
TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
FILE REFERENCE: 97-05D1
CURRENT APPLICATION NUMBER: US/09/369,617
CURRENT FILING DATE: 1999-08-06
EARLIER APPLICATION NUMBER: US 09/071,101
EARLIER FILING DATE: 1998-05-01
EARLIER APPLICATION NUMBER: US 60/044,015
EARLIER FILING DATE: 1998-05-06

NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 755
TYPE: PRT
ORGANISM: Homo sapiens
US-09-369-617-2

Query Match 11.4%; Score 63.5; DB 3; Length 755;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 22; Conservative 8; Mismatches 17; Indels 13; Gaps 4;

QY 4 LTGAGVSCGIDPFRSDGIYARLAVDPDL---PPQAMFDEIYFRKDPREFKFAKE 60
Db 147 LSGGGIVS--VLD---DGI---EKDHPDLNMYDPLASYDNDYDPDPQPRYSKE 196

RESULT 7
US-08-447-010-8
Sequence 8, Application US/08447010
Patent No. 5770718
GENERAL INFORMATION:
APPLICANT: MOFEAT, BARBARA
TITLE OF INVENTION: GENE FOR APRT FROM PLANT TISSUE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 330 University Avenue, Suite 701
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,010
FILING DATE: 22-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/230,695
FILING DATE: 21-APR-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,132
FILING DATE: 26-MAY-1992
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1811-183 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
TELEX: 065-24567 SIMBAS
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-447-010-8

Query Match 10.8%; Score 60; DB 1; Length 180;
Best Local Similarity 29.4%; Pred. No. 5.7;
Matches 25; Conservative 6; Mismatches 30; Indels 24; Gaps 4;
QY 24 IYARLAVDPDLPPQAMF-DIEYFRKDPREF-----EFKAEIYFGQ----- 65

Db 8 LVARRIVFPPPIGVLFRRDPSPLKPDSPFRASIRLASHLSTHSGKIDYIAGLDSR 67
 Oy 66 ---FOPSLCHKF---IALSDKEGKL 84
 Db 68 GFLFSPSLAQELGVGCVLIRKQKGL 92

RESULT 8 US-08-204-656B-2

; Sequence 2, Application US/08204656B
 ; Patent No. 5538882
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsui, Ikuo
 ; APPLICANT: Ishikawa, Kazuhiko
 ; APPLICANT: Miyairi, Sachio
 ; APPLICANT: Honda, Koichi
 ; TITLE OF INVENTION: Variant-Type Carbohydrate Hydrolase,
 ; TITLE OF INVENTION: Variant Gene Of The Enzyme And Method For Producing
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
 ; STREET: 8110 Gatehouse Road, Suite 500 East
 ; CITY: Falls Church
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22042
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/204,656B
 ; FILING DATE: 02-MAR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weiner, Marc S.
 ; REGISTRATION NUMBER: 32,181
 ; REFERENCE/DOCKET NUMBER: 234-252P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 205-8000
 ; TELEFAX: (703) 205-8050
 ; TELEEX: 248345
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 468 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-204-656B-2

Query Match 10.8%, Score 60; DB 1; Length 468;
 Best Local Similarity 21.1%; Pred. No. 21;
 Matches 27; Conservative 15; Mismatches 40; Indels 46; Gaps 5;

Oy 12 SCGIIDPFRSRD-----GIYA--RLAVD-----PDLPPQAMFDI- 44
 Db 171 SVALPDLTETSDVASVFNWVKDFVNSIDGLRIDSAKHVDGFFPDEVSPSGVSVG 230
 Oy 45 EYFRKDPFR-----PFFKAKEIYPGQFOPSLCHKFIALSDKEGKL 85
 Db 231 EYFGQDPAVYTCGYNYITGVSNYPLYPTTRFKTTDSSSSSELTOMISSVASSCDPTLL 290
 Oy 86 RNYTONID 93
 Db 291 TNEVENHD 298

RESULT 9
 US-07-857-224B-30

; Sequence 30, Application US/07857224B
 ; Patent No. 5958784
 ; GENERAL INFORMATION:
 ; APPLICANT: Benner, Steven A.
 ; TITLE OF INVENTION: Predicting Folded Structures of Proteins
 ; NUMBER OF SEQUENCES: 114
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Steven A. Benner
 ; STREET: Hadlaubstrasse 151
 ; CITY: Zurich
 ; STATE: none
 ; COUNTRY: Switzerland
 ; ZIP: (note: this is an international post code) CH-8092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 7.0
 ; SOFTWARE: Microsoft Word
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/857,224B
 ; FILING DATE: 03/25/92
 ; CLASSIFICATION: 436
 ; PRIOR APPLICATION DATA: none
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (International) 41 1 632 2830
 ; TELEFAX: (International) 41 1 262 2437
 ; TELEX: none
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 274
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; DESCRIPTION: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Schizosaccharomyces pombe
 ; FEATURE: Protein kinase; Table 8 Column 34
 ; PUBLICATION INFORMATION:
 ; AUTHORS:
 ; AUTHORS: Hanks, S. K.
 ; AUTHORS: Quinn, A. M.
 ; AUTHORS: Hunter, T.
 ; TITLE: The protein kinase family
 ; JOURNAL: Science
 ; VOLUME: 241
 ; PAGES: 42-52
 ; DATE: 1988
 ; US-07-857-224B-30

Query Match 10.7%, Score 59.5; DB 2; Length 274;
 Best Local Similarity 32.9%; Pred. No. 12;
 Matches 23; Conservative 4; Mismatches 16; Indels 27; Gaps 4;

Oy 50 DPEPFKAKEIYPGQFOPSLCH--KFI-----ALSDKEGKL-----L 85
 Db 104 DPLVQKFTYQVNG---VNFCHSRRIITHRLKPNLLIDKRGNIKLADFLARSGVPL 160
 Oy 86 RNYTONIDTL 95
 Db 161 RNYTHEIVTL 170

RESULT 10
 US-08-318-947A-19
 ; Sequence 19, Application US/08318947A
 ; Patent No. 5798245
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, Paul J.
 ; APPLICANT: Tian, Qingsheng
 ; TITLE OF INVENTION: TTA-1 BINDING PROTEINS AND ISOLATED
 ; NUMBER OF SEQUENCES: 21

FILING DATE: 04-FEB-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/318,947
 FILING DATE: 06-OCT-1994
 APPLICATION NUMBER: 08/133,530
 FILING DATE: 07-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mack, Susan J.
 REGISTRATION NUMBER: 30,951
 REFERENCE/DOCKET NUMBER: A462
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)293-7060
 TELEFAX: (202)293-2920
 TELEX: 6451103
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 282 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-795-303-19

Query Match 10.7%: Score 59.5; DB 2; Length 282;
 Best Local Similarity 32.9%; Pred. No. 12;
 Matches 23; Conservative 4; Mismatches 16; Indels 27; Gaps 4

QY 50 DDPPEFKAEKYEIPQOPSLCH--KFI-----ALSDKEGKL-----L 85
 ||| || :: || || || || || ||
 Db 107 DPRLVQKTKTYQLVNG---VNFCHSRRIIRDLKPONLLIDEGNKLADFGLARSTGVPL 163
 |||| | | | | | | | | |
 QY 86 RNYTONIDTL 95
 |||| | | | | | | | | |

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-874-347-22

Query Match 10.7%; Score 59.5; DB 2; Length 297;
Best Local Similarity 32.9%; Pred. No. 13;
Matches 23; Conservative 4; Mismatches 16; Indels 27; Gaps 4;

QY 50 DRRPFKFAKEIYPOFOPSLCH--KFI-----ALSDKEGKL-----L 85
DB 107 DRLVQKFTYQLVNG---VNFCHSRRIIHRDLKPNLLIDKGNLKLADFGIARSGVPL 163
QY 86 RNYTONIDTL 95
DB 164 RNYTHEIYTL 173

RESULT 13
US-09-093-522-22
Sequence 22, Application US/09093522
Patent No. 6015700

GENERAL INFORMATION:
APPLICANT: Limper, Andrew H.
APPLICANT: Leof, Edward B.
APPLICANT: Thomas, Charles F.
APPLICANT: Gustafson, Michael P.
TITLE OF INVENTION: CDC2 PROTEIN KINASE FROM PNEUMOCYSTIS
TITLE OF INVENTION: CARINII
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C., P.A.
STREET: 60 South Sixth Street, Suite 3300
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/093.522
FILING DATE: 08-JUN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/874,347
FILING DATE: 13-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ellinger, Mark S.
REGISTRATION NUMBER: 34,812
REFERENCE/DOCKET NUMBER: 07039/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-335-5070
TELEFAX: 612-288-9696
TELEX:

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-093-522-22

Query Match 10.7%; Score 59.5; DB 3; Length 297;

Best Local Similarity 32.9%; Pred. No. 13;
Matches 23; Conservative 4; Mismatches 16; Indels 27; Gaps 4;

QY 50 DRRPFKFAKEIYPOFOPSLCH--KFI-----ALSDKEGKL-----L 85
DB 107 DRLVQKFTYQLVNG---VNFCHSRRIIHRDLKPNLLIDKGNLKLADFGIARSGVPL 163
QY 86 RNYTONIDTL 95
DB 164 RNYTHEIYTL 173

RESULT 14
US-08-419-652-7
Sequence 7, Application US/08419652
Patent No. 5831007

GENERAL INFORMATION:
APPLICANT: Chua, Anne O
APPLICANT: Gubler, Ulrich A
TITLE OF INVENTION: INTERLEUKIN-12 RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/419,652
FILING DATE: 11-APR-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/248,532
FILING DATE: 31-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/094,713
FILING DATE: 19-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kass, Alan P
REGISTRATION NUMBER: 32142
REFERENCE/DOCKET NUMBER: CD 9174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 620 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 1..620
OTHER INFORMATION: 950 of leukemia inhibitory factor-
OTHER INFORMATION: receptor..."

US-08-419-652-7

Query Match 10.7%; Score 59.5; DB 2; Length 620;
Best Local Similarity 23.8%; Pred. No. 35;
Matches 25; Conservative 16; Mismatches 33; Indels 31; Gaps 6;

QY 1 IIVLTGAGVSVCGLPDRSRDGIYARLAVDPDLPDQAMDIYFRKDRPFKFAKE 60
DB 516 VAVIVGVTSITLC---YRKREMIKETP---YPDIPDEN-----CKA 551

```
QY      61 IYPGQPSPSLCHKFIALSDKEGKLLRNYT-QNIDTLLEQVAGIQRI    104
       : ||| : | : | : | : | : | : | : | : | : | : | : |
Db     552 L--QFGKSVCEGSSALKLTLE---MNPCTPNNVVEVLERSAFPKI    590
```

RESULT 1.5
HS-07-797-

```

US-07-797-556-6
: Sequence 6, Application US/07797556
: Patent No. 5262592
:
: GENERAL INFORMATION:
: APPLICANT: Geating, David P.
: TITLE OF INVENTION: Receptor for Oncostatin M and Leukemia
: TITLE OF INVENTION: Inhibitory Factor
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/797,556
: FILING DATE: 19911122
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Seese, Kathryn A.
: REGISTRATION NUMBER: 32,172.
: REFERENCE/DOCKET NUMBER: 2607
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-587-0430
: TELEFAX: 206-587-0606
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1001 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-797-556-6

```

Query Match	10.78:	Score 59.5:	DB 1:	Length 1001:
Best Local Similarity	23.88:	Pred. No. 66:		
Matches	25:	Conservative	16:	Mismatches 33:
				Indels 31:
				Gaps 6:

```

QY      1 IIVTGGVSVSCSIPFERSDGIYARLAVDFPLDPPQAAFDIETFRKDRPPEFKA 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      846 VAVTGVVTSLC---YRKREWTKETP---YDIPNPN-----CKA 883
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 IYPOQFQPSLCHKTALSDKSGKLLRNT-ONITLQVAGIQRI 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      882 L---QFQSVGSSGSSAKTLE---MNPCTPNNVVLELTSRFPRI 920
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: February 16, 2001, 10:44:13
Job time: 155 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:49:17 ; Search time 220.15 Seconds
(without alignments)
73.262 Million cell updates/sec

Title: US-09-461-580A-19
Perfect score: 556
Sequence: 1 IIVLTGAGVSVSGIPDRS.....NYTQNDTLEQVAGRIQLQ 106

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 936741 seqs, 152156132 residues
Total number of hits satisfying chosen parameters: 936741

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents-AA:*

1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097.COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098.COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099.COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100.COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101.COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US102.COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US103.COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US104.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	556	100.0	106	US-09-461-580A-19	Sequence 19, Appl
2	556	100.0	245	US-09-461-580A-4	Sequence 4, Appl
3	556	100.0	327	US-09-461-580A-12	Sequence 12, Appl
4	556	100.0	737	US-09-461-580A-1	Sequence 1, Appl
5	556	100.0	737	US-09-461-580A-26	Sequence 26, Appl

	552	99.3	128	18	US-09-461-580A-9	Sequence 9, Appl
6	471	84.7	823	20	US-09-619-049-576	Sequence 576, App
7	471	84.7	823	22	US-60-167-217-10885	Sequence 10885, A
8	471	84.7	823	22	US-60-167-217-10885	Sequence 915, App
9	471	84.7	823	22	US-60-173-464-8827	Sequence 8827, Ap
10	471	84.7	823	22	US-60-191-637-10901	Sequence 10901, A
11	471	84.7	823	22	US-60-191-637-10901	Sequence 8549, Ap
12	471	84.7	823	22	US-60-191-637-10901	Sequence 44940, A
13	471	84.7	823	22	US-60-191-637-10901	Sequence 40, Appl
14	471	84.7	823	22	US-60-191-637-10901	Sequence 35, Appl
15	471	84.7	823	22	US-60-191-637-10901	Sequence 517, App
16	471	84.7	823	22	US-60-191-637-10901	Sequence 28899, A
17	471	84.7	823	22	US-60-191-637-10901	Sequence 10557, A
18	471	84.7	823	22	US-60-191-637-10901	Sequence 8533, Ap
19	471	84.7	823	22	US-60-191-637-10901	Sequence 16, Appl
20	471	84.7	823	22	US-60-191-637-10901	Sequence 19606, A
21	471	84.7	823	22	US-60-191-637-10901	Sequence 19606, A
22	471	84.7	823	22	US-60-191-637-10901	Sequence 9, Appl
23	471	84.7	823	22	US-60-191-637-10901	Sequence 17, Appl
24	471	84.7	823	22	US-60-191-637-10901	Sequence 189, App
25	471	84.7	823	22	US-60-191-637-10901	Sequence 189, App
26	471	84.7	823	22	US-60-191-637-10901	Sequence 15, Appl
27	471	84.7	823	22	US-60-191-637-10901	Sequence 14, Appl
28	471	84.7	823	22	US-60-191-637-10901	Sequence 11, Appl
29	471	84.7	823	22	US-60-191-637-10901	Sequence 57, Appl
30	471	84.7	823	22	US-60-191-637-10901	Sequence 20918, A
31	471	84.7	823	22	US-60-191-637-10901	Sequence 10, Appl
32	471	84.7	823	22	US-60-191-637-10901	Sequence 19599, A
33	471	84.7	823	22	US-60-191-637-10901	Sequence 19599, A
34	471	84.7	823	22	US-60-191-637-10901	Sequence 39338, A
35	471	84.7	823	22	US-60-191-637-10901	Sequence 17, Appl
36	471	84.7	823	22	US-60-191-637-10901	Sequence 21, Appl
37	471	84.7	823	22	US-60-191-637-10901	Sequence 19601, A
38	471	84.7	823	22	US-60-191-637-10901	Sequence 19601, A
39	471	84.7	823	22	US-60-191-637-10901	Sequence 6406, A
40	471	84.7	823	22	US-60-191-637-10901	
41	471	84.7	823	22	US-60-191-637-10901	
42	471	84.7	823	22	US-60-191-637-10901	
43	471	84.7	823	22	US-60-191-637-10901	
44	471	84.7	823	22	US-60-191-637-10901	
45	471	84.7	823	22	US-60-191-637-10901	

ALIGNMENTS

US-09-461-580A-19
Sequence 19, Application US/09461580A
GENERAL INFORMATION:
APPLICANT: Gaerente, Leonard
APPLICANT: Imal, Shin-Ichiro
TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
FILE REFERENCE: 0050.1618-000
CURRENT APPLICATION NUMBER: US/09/461, 580A
CURRENT FILING DATE: 1999-12-15
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 19
LENGTH: 106
TYPE: PRT
ORGANISM: Mus musculus
US-09-461-580A-19

Query Match 100.0%; Score 556; DB 18; Length 106;
Best Local Similarity 100.0%; Pred. No. 8.4e-63;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IIVLTGAGVSVSGIPDRSRIARLAVDFPLDPOAMPDIEYRKDPREFRAKE 60
|||||

Db 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
QY 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106
Db 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106

RESULT 2
US-09-461-580A-4
; Sequence 4, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050,1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-4

Query Match 100.0%; Score 556; DB 18; Length 245;
Best Local Similarity 100.0%; Pred. No. 2,7e-62;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
Db 33 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 92
QY 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106
Db 93 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 138

RESULT 3
US-09-461-580A-12
; Sequence 12, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050,1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-12

Query Match 100.0%; Score 556; DB 18; Length 327;
Best Local Similarity 100.0%; Pred. No. 4,1e-62;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
Db 33 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 92
QY 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106

Db 93 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 138

RESULT 4
US-09-461-580A-1
; Sequence 1, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050,1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-1

Query Match 100.0%; Score 556; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 1,3e-61;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
Db 248 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 307
QY 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106
Db 308 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 353

RESULT 5
US-09-461-580A-26
; Sequence 26, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050,1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-26

Query Match 100.0%; Score 556; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 1,3e-61;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
Db 248 IIVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 307
QY 61 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 106
Db 308 IYPGQFOPSLCHKFTALSDKEGKLLRNTONTIDTLEOVAGIQRILQ 353

RESULT 6
US-09-461-580a-9
; Sequence 9, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580a-9

Query Match 99.3%; Score 552; DB 18; Length 128;
Best Local Similarity 100.0%; Pred. No. 3.6e-62;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 61
Db 1 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 60
OY 62 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 106
Db 61 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 105

RESULT 7
US-09-619-049-576
; Sequence 576, Application US/09619049
; GENERAL INFORMATION:
; APPLICANT: YANDELL, MARK
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSECTICIDAL
; TITLE OF INVENTION: TARGETS
; FILE REFERENCE: CL000735
; CURRENT APPLICATION NUMBER: US/09/619,049
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/171,590
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/171,627
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/175,763
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/175,685
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/186,663
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 60/187,241
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 1533
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 576
; LENGTH: 823
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-619-049-576

Query Match 84.7%; Score 471; DB 20; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50;
Matches 85; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

OY 1 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 60
Db 224 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 283
OY 61 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 106
Db 284 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 329

RESULT 8
US-60-167-217-10885
; Sequence 10885, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10885
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-10885

Query Match 84.7%; Score 471; DB 22; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50;
Matches 85; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

OY 1 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 60
Db 224 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 283
OY 61 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 106
Db 284 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 329

RESULT 9
US-60-171-627-915
; Sequence 915, Application US/60171627
; GENERAL INFORMATION:
; APPLICANT: Yandell, Mark
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000177
; CURRENT APPLICATION NUMBER: US/60/171,627
; CURRENT FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-171-627-915

Query Match 84.7%; Score 471; DB 22; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50;
Matches 85; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

OY 1 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 60
Db 224 IYLTGAGVSVSCGIPDFSRSDGIYARLAVDPDLPDQAMFDIEYFKRDPPEFKFAKEI 283
OY 61 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 106
Db 284 IYPOFQPSLCHKFTALSDKEGKLLRNTONTIDLEAVAGIQRILQ 329

RESULT 10
US-60-173-464-8827
; Sequence 8827, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: C1000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8827
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-8827

Query Match 84.7%; Score 471; DB 22; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50; 7; Indels 0; Gaps 0;
Matches 85; Conservative 14; Mismatches 7;

OY 1 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 60
|||||
DB 224 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 283
|||||
OY 61 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORIIQ 106
|||||
DB 284 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORVIE 329
|||||

RESULT 11
US-60-191-637-10901
; Sequence 10901, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: C1000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10901
; LENGTH: 823
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-637-10901

Query Match 84.7%; Score 471; DB 22; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50; 7; Indels 0; Gaps 0;
Matches 85; Conservative 14; Mismatches 7;

OY 1 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 60
|||||
DB 224 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 283
|||||
OY 61 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORIIQ 106
|||||
DB 284 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORVIE 329
|||||

RESULT 12
US-60-191-681-8549
; Sequence 8549, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.

; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8549
; LENGTH: 823
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-681-8549

Query Match 84.7%; Score 471; DB 22; Length 823;
Best Local Similarity 80.2%; Pred. No. 1e-50; 7; Indels 0; Gaps 0;
Matches 85; Conservative 14; Mismatches 7;

OY 1 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 60
|||||
DB 224 IIVLTGAGVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKE 283
|||||
OY 61 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORIIQ 106
|||||
DB 284 IYPGQFQPSLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORVIE 329
|||||

RESULT 13
US-09-270-767-44940
; Sequence 44940, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 44940
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44940

Query Match 78.1%; Score 434; DB 16; Length 286;
Best Local Similarity 78.6%; Pred. No. 1.2e-46; 7; Indels 0; Gaps 0;
Matches 77; Conservative 14; Mismatches 7;

OY 9 VVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKEIYGCYOP 68
|||||
DB 1 VVSVCIGIPDRSRDGIYARLAVDPDLPDQAMFDIEYFRKDRPFYKFAKEIYGCYOP 60
|||||
OY 69 SLCHKFTALSDKEGKLLRNTONTIDLEOVAGIORIIQ 106
|||||
DB 61 SPCHRFKMETKGLRNTONTIDLEOVAGIORVIE 98
|||||

RESULT 14
US-09-461-580a-20
; Sequence 20, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichi
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A

```

:
: CURRENT FILING DATE: 1999-12-15
:
: NUMBER OF SEQ ID NOS: 35
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO: 20
:
: LENGTH: 107
:
: TYPE: prt
:
: ORGANISM: Mus musculus
:
: US-09-461-580A-20

```

Query Match	56.2%	Score 312.5	DB 18	Length 107
Best Local Similarity	59.2%	Pred. No. 9e-32		
Matches 61	Conservative 17	Mismatches 22	Indels 3	Gaps 2

OY 1 IIVLTGAGVSVSCGIPDERS-RDCIYYARLAVDFPLEDDPQAMEDIYEKKDPREFEAK 59
:
| | : | | | | | | : | | : | | | |
Dd 1 VICLVGAGISYSAGIPDRSPSTGLYANL--EKYHLYPEALFEISFYKHPPEPFALAK 58

```
QY 60 EIIYGOPPSLCHKETALSDEKGLIRNTQNIDTLEAVAGIQ 102
|::|||::|| | :: || |||||::||::
Db 59 ELVGGQFPPTICHYFIRLLTKKEGILLRCYTQNIDTLEAVAGIE 101
```

```

RESULT 15
US-09-461-580A-35
: Sequence 35, Application US/09461580A
: GENERAL INFORMATION:
: APPLICANT: Guarente, Leonard
: APPLICANT: Imai, Shin-ichiro
: APPLICANT: Armstrong, Christopher
: TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
: TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
: TITLE OF INVENTION: LIFESPAN
: FILE REFERENCE: 0050.1618-000
: CURRENT APPLICATION NUMBER: US/09/461,580A
: CURRENT FILING DATE: 1999-12-15
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 35
: LENGTH: 232
: TYPE: PRT
: ORGANISM: unknown
: FEATURE:
: OTHER INFORMATION: Deduced Amino Acid Sequence of cDNA
: US-09-461-580A-35

```

Query Match	56.2%	Score 312.5	DB 18	Length 232
Best Local Similarity	59.2%	Pred. No. 2.7e-31		
Matches 61	Conservative 17	Mismatches 22	Indels 3	Gaps 2

QY 1 IIVITGAGSVSCGIDPFRS-RDGIYARLAVDPEDLPDPOAMFIEYERKCDPRFFEFKA 59
:
87 VICLVGAGISTSAGIPDFRSPSTGLYANL-EKYHLPYPALFEIYSFKKHDEPFALAK 144

Dy 60 EIYVGQEPRLCHKFTALSDKEGKLRYNTQNIDTLEVAQIO 102
| : | | | | : | | : : | | | | | | | | :
Dd 145 EIVYGQFRTICHYFRLLKREGGLLRCTQNIDTLEVAGIE 187

Search completed: February 16, 2001, 10:49:18
Job time: 459 sec

22	96.4	2.5	1856	10	AF095714	Homo sapi
23	96.4	2.5	1859	17	AK025876	Homo sapi
24	96.2	2.5	1845	48	AX017253	Sequence
25	94.8	2.5	2393	47	SC1D042C	Z74090 S.cerevisia
26	94.8	2.5	4649	47	SC5IR2	X01419 Yeast SIR2
27	94.2	2.5	36687	47	SCCIVL37K	Z71819 S.cerevisia
28	94.2	2.4	67200	2	MTV017	AL021897 Mycobacte
29	92.8	2.4	87283	50	AC022856	Homo sapi
30	92.6	2.4	49200	50	AC022749	AC022856 Homo sapi
31	91.2	2.4	171574	39	AC012300	AC012300 Homo sapi
32	91.2	2.4	63523	51	AC024266	AC024266 Homo sapi
33	90.4	2.3	840	7	CNS01BNS	AL114464 Botrytis
34	90.4	2.3	141892	50	AC023197	Mus musc
35	90.2	2.3	1576	10	AF131800	AF131800 Homo sapi
36	90.2	2.3	137431	38	AC010214	AC010214 Drosophil
37	89.8	2.3	80133	49	AC021347	AC021347 Homo sapi
38	89.6	2.3	200580	66	AC084064	AC084064 Homo sapi
39	89.2	2.3	78220	50	AC023212	AC023212 Homo sapi
40	89	2.3	4492	11	RAT55ANTE	D37934 Rat mRNA fo
41	89	2.3	76052	50	AC023262	AC023262 Homo sapi
42	88.8	2.3	36082	50	AC022663	AC022663 Homo sapi
43	88.4	2.3	62649	50	AC022552	AC022552 Homo sapi
44	88.4	2.3	220469	59	AC074307	AC074307 Mus muscu
45	87.6	2.3	82615	55	AC062001	AC062001 Homo sapi

ALIGNMENTS

BASE COUNT	1118 a	761 c	889 g	1081 t
ORIGIN				
	VTDPFDLDPQAMQFDIEYFRKDPREFEFKACTIKYGGOPSLCHFEIALDSKEKCLLANN YTKQNDLPDEQVAPQOIRILIQCHGSFATSLCKKRYVCEANRGIOIFNOVYIPRCEGRRA DEPLAIKMKPELVFFGENDLPDFHRAMKIDKKEVDLLVIGSSIKVRVVALIPSIPIHEE VPQLILINKEPLPHLFLVELLGGDCVILINELHGGEGVYALNCNPKRSLSEIKPR POKELVHLSELPPPLHLSEDSSESPRIVPODSSVIAVLDOATNNVNDLEVESSC VEEKPOEVOYTSRNVENINVENPDERKAVGSTRADKERTSVATYRKCMPRMLAEQIJS KRLGNQDLPVPMRYRITFHGAEEVYSDSDYVLSSSGSSNSDCTCOSPSEPLEEDD SEIEEFYNGLEDLDERPEKACGSGFGGDGGQEVENVAINTROELTVNVNPSDKS"			
Query Match	99.5%; Score 3849; DB 11; Length 3849;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 3849; Conservative	0;	Mismatches	0;	Indels
	0;	Gaps	0;	
QY	1	gcgagacgaagagagcgagggcgagagggccagagaagcgagtgtgaagaatggcgagagag	60	
Db	1	CGGGAGCAGAGAGAGCGGAGGCGCGAGCGCCAGAGCGCAGTTCGAACATGCGCGAGCAGAG	60	
QY	61	tggcgctgcgccttaagacgcgcgcgcgtccctctcgagggcgagcgcccatgtgaagcgcgct	120	
Db	61	TGGCGCTGCCTTCAAGGCGCGGCGCTCCCTTCCGCGGCGGCGCCGCAATGGAGGCGGCGCT	120	

RESULT	1
LOCUS	AF214646 3849 bp mRNA ROD 15-JAN-2000
DEFINITION	Mus musculus Sir2alpha protein (Sir2alpha) mRNA, complete cds.
ACCESSION	AF214646
VERSION	AF214646.1 GI:6693710
KEYWORDS	.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 3849)
AUTHORS	Imai,S., Armstrong,C.M. and Guarente,L.
TITLE	Silencing and ageing protein Sir2 is an NAD-dependent histone deacetylase
JOURNAL	Unpublished
RESEARCH	2 (bases 1 to 3849)
AUTHORS	Imai,S., Armstrong,C.M. and Guarente,L.
TITLE	Direct Submission
JOURNAL	Submitted (10-DEC-1999) Dept. of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
FEATURES	Location/Qualifiers
SOURCE	1..3849
gene	/organism="Mus musculus" /strain="Swiss Webster/NIH" /db_xref="taxon:10090" /dev_stage="15-day embryo" 1..3849
CDS	/gene="Sir2alpha" 48..2261 /feature="Sir2alpha" /function="have NAD-dependent histone deacetylase and putative mono-ADP-riboseyltransferase activities" /note="Closest homolog to the yeast silencing and aging protein, Sir2p" /codon_start=1 /product="Sir2alpha protein" /protein_id="AAF24983.1" /db_xref="GI:6693711" /translation="MADEVLALQAGSPSAAMEAASOPADEPLRPRPDGGL RSPEPSAAVAAPAAAGCAASAAPALMPREAAGAASAEKPAATVAVGGDNGSG RRERADDPPDDGEEDDEAAAAAAAATCGRNLLITDGLTNGFSCSDDDR SHASSDMTPPRIGPTTVVOOHLMICTDPPTIIKDLLEPIIPPELDWTLWOIVTI ILTSPPRKRRKKDINTIEDAVKLLOEQEKIIIVLGAGVSVCGLPDFSRGIRYARL

OY	181	gcaagccgagcgagccgagcgcaacagatgagcgcccgcgcgcggtgtgaaagcgcgga	24.0
Db	181	gcagcccgccggcgagccgagccacagatgagggccgccgccgccgggtgtgagccggcgga	24.0
OY	241	gcgccgcgagcccgcgagcgagctgtgtgacggagagcgcaagggcgcgcgagcgcgagc	30.0
Db	241	ggcgcccgccggcccgccggcgagctgtgtgacggagagcgagggcgccggcgagcgcgagc	30.0
OY	301	ggagagcccgcgagcgagcgtgtgcggggagcgagagcaatgggtccgagctgcggggg	36.0
Db	301	ggagagcccgcgagcgagcgtgtgcggggagcgagagcaatgggtccgagctgcggggg	36.0
OY	361	agccgagggcgagcttgaagacttcgacgagcgagggcgagagagagagagcgagcgcg	42.0
Db	361	agccgagggcgagcttgaagacttcgacgagcgagggcgagagagagagagcgagcgcg	42.0
OY	421	cgagcagcgcgagcgagcgagatcgagctaccgagagcaacctctgtgacgagatgactcc	48.0
Db	421	cgagcagcgcgagcgagcgagatcgagctaccgagagcaacctctgtgacgagatgactcc	48.0
OY	481	tcaactaagagcttcatctccgtgtgaaagttga ttgacga ttgacagagcttcaacgagct	54.0
Db	481	tcacactaagagcttcatctccgtgtgaaagttga ttgacga ttgacagagcttcaacgagct	54.0
OY	541	ctagtgacttgagctccgagcgcgcggaaggtccataactttgtcaagaacatctca	60.0
Db	541	ctagtgacttgagctccgagcgcgcggaaggtccataactttgtcaagaacatctca	60.0
OY	601	tgattggagacgagatccctcgaaacaattttiaaagtttatatccgagaaacaattccctcac	66.0
Db	601	tgattggagacgagatccctcgaaacaattttiaaagtttatatccgagaaacaattccctcac	66.0
OY	661	ctgagcttgagatcgatatgagctctgagcagattgtcttaataatcccttcagaaaccaaa	72.0
Db	661	ctgagcttgagatcgatatgagctctgagcagattgtcttaataatcccttcagaaaccaaa	72.0
OY	721	agcgagaaaaaagaagaataatcaataatgaagatgctgtgaaagtacatcgagagagt	78.0
Db	721	agcgagaaaaaagaagaagatgtcaatgtcaattgaattgaatgagctgtgaattctcagagagt	78.0
OY	781	gtaaaaaagataagcttcgagctgagacgtgggtttcttcctctctgaggatctctgact	84.0
Db	781	gtaaaaaagatataagttctgacttgagagctggggtttcttcctctctgaggatctctgact	84.0
OY	841	tcagatcaagaagacggtatcatctgactgcgcttgggtgagacttccagacatcccaagc	90.0

Db	841	TCGATTCAGAGAGCGGATATATAGCTCGCCTTGCGGTGGACTTCCCAACTCCAGACC	900
Qy	901	ctcaagccatggttgbatattgagtaatttagaanaagaccagaaccattcttcaagttg	960
Db	901	CTCAAGCCATGTTGATATATGAGTATTTTGAAGAAGACCAGCAACATTCTTCAAGTTTG	960
Qy	961	caaaaggaatatatacccggaacagttccagccgtctctgtgtcaaatatcaatgttct	1020
Db	961	CAAAAGGAATATATTCGGGACAGTTCCAGCCGCTCTGTGTCCAAATTCATACGTTTGT	1020
Qy	1021	cagaataagagaagaaataactcttcgaataataaccaaatatagataaccttggaagcg	1080
Db	1021	CAGATTAAGGAAGAAACTTACTTCGAAATTAATCTCAAAATTAATATCTTTGAGCGG	1080
Qy	1081	ctgcaaggaatcccaagagatccctcgaigtgcacagtgctcccttgcacagcatcttgccgta	1140
Db	1081	TTGCAAGGAATCCAAAGATTCCTTCACTGTCAATGTTCTCTTGCAACACATCTTGCCGTA	1140
Qy	1141	ctgtgtaataatacaagtgtatgttgaagcgtgtctgtgtagaacaatttcaacagtgatgc	1200
Db	1141	TTTTGTAATATCAAAAGTGTGATTTGTAAGCGTTTCGTGGAGACATTTTATACGAGTAGTTC	1200
Qy	1201	ctcggtgccttagtgcccaagcgtgatggaacacacttgcacataatgaagccaagattgtct	1260
Db	1201	CTCGGTGCTCCCTAGTGCGCCAGCTGATGAGACCCATCTGCATCTGGAAGCCAAAGATTGCT	1260
Qy	1261	tctcttggtagaaacttaccagaacagttccatagagccatgaaatgaaatgaaatgaag	1320
Db	1261	TCTTTGGTGAAGAACTTATCCGAGAACATTTCTATAGACCCTGAAGATATATACAAAGTAGAG	1320
Qy	1321	ctgaacctccatctgttatgtgatactctctctgaaagtgaagacagtagaactaatccaa	1380
Db	1321	TTGACCTCCATCTGATTTGATATGAGATCTTCTGTAAGTAGAGACCATGACATTAATTCOA	1380
Qy	1381	gttctataaccccatgaaatggtcccaatatataataatagggaaaccttgcctcatctac	1440
Db	1381	GTTCATAACCCCATGAAGGTCCCTCAAAATTTAATAAATGAGAACCTTTGGCTCATCTAC	1440
Qy	1441	attcttgaatgtagagctcccttggaagctctgagatgattaaataatgagtgtgcataagc	1500
Db	1441	ATTTTGATGTAGAGCTCTCTTGAGAGACTGGGAGTTRTAATTAATGATGTGTCTATAGCC	1500
Qy	1501	taggtgtagaatalgtccaaacttgttctgaacctgtlaaacctgttcaagaaatlaactgaaa	1560
Db	1501	TAGTGTGATATATGCCAAACTTTGTTGTATACCCGTATAAGCTTTCAGAAATTTACTGAAA	1560
Qy	1561	aaactccagcccaaaaaggaattggttcaattatcagagttgcaccaaacactctctc	1620
Db	1561	AACCTCACGCCCCANAAAGAAATTTGTTCAATTTATCAAGTTTGCACCAACACTCTTTC	1620
Qy	1621	atattcttggaagagctaaagttcaacctggaagaagactgtaccaaaagacctctgttgatg	1680
Db	1621	ATATTTTGGAGAGCTCAAGTTCACTGTGAAGAAGCTGTACCAAAAGACTCTTCTGTGATGG	1680
Qy	1681	ctaacactgttagaccgaagcacacaacaacaatgattaaatgatttagaagtatctgaatcaa	1740
Db	1681	CTAACCTTTGTAGACCAAGCACAAACAAACATGTTATATGATTTAAGATATCTGAATCAA	1740
Qy	1741	gttgtgtgtagaagaaaaaaccaagaagtacagactagtagaagtgtgagaacatlaatg	1800
Db	1741	GTGTGTGTGGAAGAAAAACCAAGAGAGTACAGACTATAGTAGAATTTGTGAGAACTTTAATG	1800
Qy	1801	tgggaataatccagaattttaagcgctgtgtgttcagtagtgcagacacaaaatgaagaactt	1860
Db	1801	TGGAAAAATCCAGATTTTAAGCGCTGTGTGTTCAGATATGTGCACACAAAAATCAAAGACTTT	1860
Qy	1861	cagtttgagaagaaacagttagaanaatgtctggccctaataagacttgcagaagagcagattga	1920
Db	1861	CAGTTTGCAGAAACAGTACGAAAAATGCTGGCCTATATAGATTTCAGAAAGACAGATTAGTA	1920
Qy	1921	aggcgcttgagggtaatcaatatacctgttgttaaccacaaatcgttataatctcaagctgt	1980

Db	1921	AGCGGCTTGAGAGGTAATCAATTAACCTGTTTGTAACCAACCAATGTTACATATTCCACGGTG	1900
QY	1981	ctgaggtatactacgactcgtgaagatgagcgtctgtcctctagttcctctgtgtagcagtaaca	2040
Db	1981	CTGAGGTATACTCGACTCTGGAAGATGAGCGTCTTGCTCTAGTTCTCTGGACGATACA	2040
QY	2041	gtgacagtgagacatgcgcaagatcttgaaagaaccccttgtaagaaatgaagtgaa	2100
Db	2041	GTGACAGTGGCACATGCGCAAGTCCAAAGTTTAAACAAACCTTGGAAATGAAAGTGAA	2100
QY	2101	ttgaaagaattctaaacggtcttgaaagatgatacggagagagcgcaatgtgctgaagat	2160
Db	2101	TTGAGAAGATTCTCAATATGCTTTGGAAGATGATACGAGAGGCCCAATGTCTGGAGAT	2160
QY	2161	ctggattctggagcgtgagtgagggagatcaagaggtgttctaatgaagctaagctacaagc	2220
Db	2221	AGGAATTGACAGATGTAAACTATTCATCCATGACAAATCATACATTAATGAAGCTGTCGG	2280
QY	2281	atccaagaattgctccacgcaatggtgggaacttagcatgataaaaataatgaatgttac	2340
Db	2281	ATTGAGGAATGCTCCACGACGATGGGAACCTTACCATGCTCAAAAATAATGAATCTTTAC	2340
QY	2341	ctggaacttgaaacaagaatctgaaagatgatatattctatagactgaaataagatg	2400
Db	2341	TTTGGAACCTTCAACAGGAATCTGAAGAATGATATTATTATAGACTGGAAATATGATTG	2400
QY	2401	tctctctggaaattcttaaaagttccatcatctctgtttgtactgtgcatctcaacacg	2460
Db	2401	TCTTCTGGAAATTTCTTAAAGTTCCATCATCTTCTGTGTGACTGTGATCTCAACACTG	2460
QY	2461	tttgttgaactcatctcctcttcaagttcatcttgatgatatcatctgtatgatatgata	2520
Db	2461	TTGGTGACTTCATCTTCTTAAAGTTCATTTGATATATACATTCGATGTATATGATTA	2520
QY	2521	atttgtttcttgcttaatgatttcaacctttaaagtttcaaagcattgtaagtg	2580
Db	2521	ATTTTGTTTTTGGCTAATATAGTTTCAACTTTTAAAGTTTCAAAAGCCATGGAATGT	2580
QY	2581	taatgtaaagggaaacagctatctctagacaanaaagatgatatctcaacttttctgt	2640
Db	2581	TAAATGTAAGGAGAACACTTATCTAGACCAAAAGATGTATTCACACTTTTGTGT	2640
QY	2641	aacattgtaatgtttaaagccctcaattctgtctgcgcgaaactttatttttagagac	2700
Db	2641	AACATTGATGTATTAAAGCCCTCAATTTCTGTCTGCTGAACCTTTTATTTTATAGACAG	2700
QY	2701	ttaactttttaaacaacgagatcttccaacacttggagagtaactttttaaatacaca	2760
Db	2701	TTAACTTTTAAACACGCGCATTTTCCAAACCTTGCGAGCTTAACCTTTTAAATTCACA	2760
QY	2761	gatactctgtaatgttgagggatcagaacggtgctgtgagcactccaacacttggctcag	2820
Db	2761	GATGACTTGTAATGTGAGGAGTCAAGCAGCGTGCTGAGACACTCAAAACTTGGCTCACT	2820
QY	2821	gtgtgaagcgtactactcatcgcgttttgttactgtgcgcagacgvgbtaatgtccaaac	2880
Db	2821	GTGTGAAGCGTAACTTACGTGATGTGTTTGTACTTCTCTCAACCGTGTAATGTCCAAC	2880
QY	2881	aggcccccggagaactaacctgtataatgatttggaaaatgtgttcagttggtctgtaaaa	2940
Db	2881	AGGCCCTCGAACAATCTATTAATGATTGGAAAATGTGTTCAAGTTGTTCTGAAACA	2940
QY	2941	atagtgcctgtcctaataaggtccccctagtttgaatatttgcattgttctaataaac	3000
Db	2941	ATAAGCGCTGTCTAATATAGTCCCTTAAGTTGAAATATTTGGCAATGTTTAAATTAATAC	3000
QY	3001	ctatacactgtgtgtagagcctgtgcatagactcttcaaccaaaatactgcgaagaatgtgaat	3060
Db	3001	CTATACACTGTGTGAGACCTGTGATATATGTTCAACCAAAATCTGCCAAGATGTGAATAT	3060

Db	249	GGCTGCCCCGGTGGCGGCGGCGGCGGCTGTGGCCGGAGACCGGAGCAGAGCGCGGCG	308
Qy	288	gagagcgcgagcgagcgagcgcccgcgagcgccgttgcgcggagcgagacaattggtctc	347
Db	309	GCAGGCGGGGCGAAGAGGCCCAAGCGACTGCGGGCGGCTGGGGAAAGAGACAAATGGCGG	368
Qy	348	ggctctgcyg-----cgagagcgagggcggtctgacgacttcgaacgaacgaaggc	398
Db	369	GGCGTCGAGGGCCCATCTCCGGGAGCCACCGCTGGCCGACAACTTTAGACAGCAAGACGAC	428
Qy	399	gagagagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc	458
Db	429	GACGACGAGGGCGAGGAGGAGGAAGCGCGCGCGCGCGCATTTGGTACCGGAGATTAAC	488
Qy	459	ctccgtgtgacccgaatgagacccctcactaatggtcttcactctcgttgaagtgtacgat	518
Db	489	CTTCTGTTCCGCTGATGAATAATTATCACAATAGTTTCATTCTCTGGGAAGATCATAGAG	548
Qy	519	gacagaaacgtlcaacagcgcaagcctctagtgagtcctcgcgcgcgagtagtccatat	578
Db	549	GATGAGAGCTCACAATGCACAACTCTAGAGTGTGACATCCAAAGCCACGGTAGTGCATAT	608
Qy	579	aactttgttgaagaaactlcaatctgattttgacacgataccctcgaaacattcttaagatta	638
Db	609	ACTTTTGTTCGCAACAATCTTATGATTTGGACACGATCTCGAACAATCTTAAACATTTA	668
Qy	639	ttacacgaagaacatctctccacacctgagcgtgagatagatagcgttggagatgttatt	698
Db	669	TTGGCGGAACAATACCTCCACCTGATGATGATGATATACACTGTGGCAGATTCTTATT	728
Qy	699	aatatccttltcagaaaccaccaaagcggaaaaaaaagaataatacaatacaattgaagat	758
Db	729	AATATCTCTTCAGAACACCAACCAAAAGGAAAAAAGAAAGATTAATTAATCAATGAAGAT	788
Qy	759	gctgtgaagtctacgcgcagcggtgtlaaaagataatgtctgcactcgagcctggggtcttc	818
Db	789	GCTGTGAATTAATCTGCAAGAGGTGCAAAAAATTAATGTTCTAACTGAGACCTGGGCTGTCT	848
Qy	819	gtctccgttggagatctctgacttcacgaatcaagaagacggtatcatatgctgcgcttgcgtg	878
Db	849	GTTTTCATGTGGAAATACCTGTGACTTCAGGTCAAGGAGATGATTTATGCTGCGCTTCTCTA	908
Qy	879	gacttcccaagacctcccaagacccctcaagccatgttltgatatglatlttgaagaaagac	938
Db	909	GACTTCCAGATTTTCCAGATTCCTCAAGCATGTTGATTAATTAATTCAGAAAGAT	968
Qy	939	ccaagacacatcttccaaggttctgaagaagaaatatatcccgacagttccagcgctctg	998
Db	969	CCAAGACCATTTCTCAAGTTTGGAAAGGAATTAATTCGCGCAATTTCCAGCATCTCTC	1028
Qy	999	tgtaacaattcatagcttgttgcagatlaagaagaagaaactacttcgaaattatatactaa	1058
Db	1029	TGTCACAAAATTCATAGCTTGTGCAGATTAAGGAAGGAAAACTTTCGCAACTATACCCAG	1088
Qy	1059	aataatagataccttggcgacggttgcagaaatccaaagaaatccctcaagtgtcatgtgtcc	1118
Db	1089	AACATAGACACGCTGGGAACAGGTTGGCGGGAATCCAAGAGTAATTAATCAGTCATGCTTC	1148
Qy	1119	tttgcgaacagcatcttgcctgatttgtlaaatatacaaaagtgtgatgtgaagcgttctgga	1178
Db	1149	TTTGGCAACAGCATTTCTCCGATTTTGAATTAACAAAAGTTGACGTGGAACCTGTACAGGA	1208
Qy	1179	gacatttctaactaggtgagttccctcggtgcggtccagggtgcacagcgtgataagacattgc	1238
Db	1209	GATATTTTAAATCAGTAGTTCTCTCCATGTGCTTAAGTGGCCCAAGCTGATGAACCGCTTCT	1268
Qy	1239	atcatgaagccagagatgtgtcttcttgttgaanaactlcaacgaacagttcatatagacc	1298
Db	1269	ATCATGAACACAGAGATTTGTTTTTTGGTGAATAATTAACGAACAGTTTATATAGACC	1328
Qy	1299	atgaagatatacaaaagatgaagtgtgacccctcatatgttattatgtgactctctcgaagtg	1358
Db	1329	ATGAAGATATGCAAAAGATGAAGTTGACCTTCCTATGTTATTTGGGCTTCTTCCCTCAAACTA	1388

QY	1339	agccagtagagacacaaatccaaagtctctataccccaagaagtcctccaatatataat	1418
Db	1389	AGACCAGTAGACACTAAATTCGAAGTCCATACCCCAAGAGTGCCTCAGATTAATTAAT	1448
QY	1419	agggaaccttgcctcaacttcaatttgaatgtagagctccttcggagacgcatctta	1478
Db	1449	AGAGAACCTTGGCTCATCTGCAATTTTGATGTAGACCTCTTGAGACCTGTGATCTATA	1508
QY	1479	attaatgaatgtgtgcaataagctagtgytgyaatalgccaacctgtgtgttaacctgta	1538
Db	1509	ATTATGAATATGTGTCTATAGGTTAGGGTGAGATATGCCAACTTGGCTGTAACTCTGA	1568
QY	1539	aagcttcaagaatctactgnaaaacctccagccccaagaagaattgttcaattataca	1598
Db	1569	AAGCTTTCAGAAATTAAGTGTGAAAAACCTCCACGAAACCAAAAAAGAAATGGCTTAATTTGCA	1628
QY	1599	gagttgccaccaaaacctctcataltttcggaagacataagttcaccttgaagaacttca	1658
Db	1629	GAGTTGCCACCACCACTCTTCATGTGTTGAGAGACTCAAGTTCAACGAAAGAACTTCA	1688
QY	1659	cccaagaacctctctgtgattgtctcaactgtgtgaccaagcaacaacaatgttaat	1718
Db	1689	CCACCAGATTTCTCAGGATTTGCACACTTTTGACCAAGCAGCATGAAGTAAATG--AT	1745
QY	1719	gatttagaagatctatgataca---gtgtgtggaagaanaacaaagaagtagacagt	1775
Db	1746	GATTTAGATGTGTCTGAATCAAAAGTTGTATGTGAAGAAAACACAGCAAGTACAAACT	1805
QY	1776	agtaggaatgtgtgnaacatlaatg----tggaaatccagattttaaggctgttgt	1829
Db	1806	TCTAGGAATGTGTGAAGTAATTAATGTGGCAACAGATGGAATAATCCGGATTTTGAAATGTTGGT	1865
QY	1830	tccagttactggaagcaaaaatagaagacttggttcggaacaagltgagaagaatgcgtg	1889
Db	1866	TCATGATCTGGGGGAAAAAATGAAGACTTCAGTGGCTGGAACTGTGAAAAATGCTGG	1925
QY	1890	ccataagacttgaagaagcagagatlagtaaacggtcttgaggttaatactatctgtlt	1949
Db	1926	CCTATAGAGTGGCAAGGAGAGACATTAGTAGCGGCTTGATGATATCAGTATCTGTTT	1985
QY	1950	gtaccaccaaatcgtttacatactccacggtgtctgagtagtactcagaactctgaagaatgac	2009
Db	1986	TTGGCCACCAAATCCTTACATTTTTCCAGGGCGTAGAGTATTTACAGACTCTAAGATATAC	2045
QY	2010	gtctgtcccttagttccctgttggagaagaagtagacagtaggcacatgycagagctccagt	2065
Db	2046	GCTTATTCCTTAATTCTTCTGTGGCAGTAGACAGTATAGTGGACATGCCAGAGTCCAACT	2105
QY	2070	ttagaagaaccttgsaagaatgaagaatgaatgaatgaatctcacaatgtgcttgaagaat	2125
Db	2106	TTAGAGAAACCCATGAGAGATGAAGAATGAAGAATTTACATATGCTTAAAGAAAT	2165
QY	2130	gatacggagagagcccgaaatgtcctcggagaggtctcggaatttggagctgtagatggaggaata	2189
Db	2166	GAGCTGTGTTTCCAGAGAGAGAGCTGGAGGAGCTGTGGATTTGGGACATGAGATGATATCAA	2225
QY	2190	gaggttgttaatgaagctatagctacaagaacagaaatltacacagatgtaaactatcatca	2249
Db	2226	GAGGCATTAATTAAGACGTATATCTGTGAACACGAGAAATGAACAGACATGAATATCATCA	2285
QY	2250	gacaaatataaacctatltgaagctgtccggaatcaggaatgtctccaccagatctggga	2305
Db	2286	AACAATATC--AGTGTAAATTAATTTGCGAGGTACAGGAATTTGTTCCACACCACTTAAGA	2342
QY	2310	actttagcatgtcaaaaaaagaatgttttccctgtgaac----tgaacaaagaatcgtg	2365
Db	2343	ACTTTAGATGTC---AAATGAATGTGTTACTGTGAATCATGATAGAGCAAGCAACAG	2399
QY	2366	aaagaatgtattattatagacgtg-aaatagatg-ctctcctgtgataatcttcaagt	2423
Db	2400	AAAGGTGAATATTAATTAAGTTGGTATAAAGATGTTCTTTTCATGATTAATTTAACTT	2455

Oy 2424 tccatcattctctgttac----- 2443
 Db 2460 CATTTATTTCTGACTTGTGTAACAACTCAACACTACTTTTCTTTTAAAAAAGG 2519
 Oy 2444 -----ttgtaacttaaacctgtgttgtaactcattccttccaagttcattg 2495
 Db 2520 TACTAAGTATCTCAATGACGCTGGTGGGCAAGACTACTTCTTTTAAAGGTTCAATGG 2579
 Oy 2496 tatgataactcgtatgtatgataa-----tttcttcttgcccaatgagttcaacct 2551
 Db 2580 TATGATTAATATCATATGTGTATATATATATTTTGTGTTTCTGAGTCAAGTTCAACT 2639
 Oy 2552 ttt--aaagtttcaaaagcaatggaatgtaatgta-----aagggaacagctatccta 2605
 Db 2640 TTTTAAAGTTTCAAAAGCCATCGGAATGTAAATTAATTAAGGAGGCACTAATCTA 2699
 Oy 2606 gaccagaagaatgtaattcacaacttttggttgttaacattgaaatgtaagccctca 2665
 Db 2700 GACCAAGAAATGTAATTT--TCACCTTTTCTTTGTAACATGGAATGTTTGAAGTACTCA 2756
 Oy 2666 attctctgtcgtgaactttatcttatttagagcaatgaacttttaaacacgtgacttt 2725
 Db 2757 AAATCTGTAGCTTAACCTTTGATCTTTTACACAACTTTTAAACACTGCACTTTT 2816
 Oy 2726 ccaaaactgtgcaagcaacttttaaaatcacagatgacttgaatgta-----g 2778
 Db 2817 CCNAAC-TCGTGGCAGCTTAACCTTTTAAATCTCAATATGACATGACAGTGTAGTAAGG 2875
 Oy 2779 gagtgaagacggtgtcgtgagacac----- 2804
 Db 2876 AAGTCAACAAATATGTTGGGAGAGACACTCGGTTGTCTTACTTTTAAAGTAATCTTGGT 2935
 Oy 2805 -----aaacttgggctcaggtgtgaaagcgtactactgacgttttggactgtcgt 2859
 Db 2936 GCTAAGAAATTCAGGATATGTATTTACGTTCAATGAAGATGGCTTTTGTACTCTCG 2995
 Oy 2860 cagac--gtgtaattgtccaaacagggccctgagactaactctgaataatgatt-----t 2911
 Db 2996 TGGACATGTAGTAATGTCTATTTGGCTCATTAACCTAAGCAAAATAATTAATGTCT 3055
 Oy 2912 ggaatgtgttcaatgttctcagaacaatgagtcgtctcattatagtcctcctagt 2971
 Db 3056 TTGGAATGTTTCAAGTGTCTTTAGAACATTAATGTCCTGCTGCGAT--CCCTTAGTTT 3112
 Oy 2972 tgaatattgacatgttcttaataataactcactcactgtgtagagccctgacatctt 3031
 Db 3113 TGAATATTTTGCATTTGTTTAAATACATCACTGCGTGAAGCTTTGCATTTATCTT 3172
 Oy 3032 caccacaa-----atactgccaagatgtgaatagcaaaagcccttcgaatc--taataa 3084
 Db 3173 TTCCACAAGTATTAACCTGCCAAATGTGAATATCCAAAGCCCTTCTGAATCTAATATA 3232
 Oy 3085 tggtaacttacttggggagagtgtaatttttgagactgctgttttccattaatgagaa 3144
 Db 3223 TGGTACTTCTACTGGGGAGAGTGAATATTTTGGACTGCTG-TTTTCCATTAATAGAGG 3291
 Oy 3145 agcaatagagcccttaact--taaaagccaagatgataa--taaatgtgactcaaca 3201
 Db 3292 AGCAACAGGCCCTGATTAATACAGTTCCAAAGTAAATGATTAATGATTAATGAGCA 3351
 Oy 3202 gaaatgatacgtgtgctgttgaagaaattgtgtgaatgatacccaaggltgagcctgt 3261
 Db 3352 GAAATGATCATGTCCTCCATTGGAGGATTTGGTGTAAATACCAAACTGCTAGCCCTAGT 3411
 Oy 3262 attatggagag-----atacagatccaatagtcataatgaaact 3301
 Db 3412 ATTATGGAGATGAACATGATGTAACCTTGAATAAGCAAGATAGTAAATGAATGAAC 3471
 Oy 3302 agttctta-----gtatttaaaagcttagcttcccttaaaactgaggaatatt 3352
 Db 3472 AGTCTTATTAATTTATCTTTTATTAAGCTTAGCCTGCTTAATAACTAGACATTAAC 3531
 Oy 3353 tctcaactgcagaacttttagccttccaacagttcacactca-----gaaagtc 3404

Db 3532 TCTCAGCTGCAAAAGCTTCTAGTCTTTCAAGAACTTCACTTTATGAATTCACAGTA 3591
 Oy 3405 agtattttttacagacttcttggaacatgtgccccaaatataatca-----tg 3459
 Db 3532 AGCATTTATTTTTCAGACCATTTTGAACATCAGCTCAATTAATTAAGTATTCCTCG 3651
 Oy 3460 tgggtttagtatttatacaaa-aaacagattgaaatagctgttctttagcataa 3518
 Db 3652 TTGCTTAGTATTTATTAACAATAAAGGTTTGAATATAGCTGCTTTATGATATAA 3711
 Oy 3519 ataaccagttagaccattactgcagagggagaaagtaataagta-gtcaattcccta 3577
 Db 3712 ACACCACATGAGACCATTAACCTAGCCAGAGAAAAAATGTATTAATGATTCCTCCTA 3771
 Oy 3578 cctaaagat-----aacgtattttttgtgctacactaaagaaatgcaatatttagt 3631
 Db 3772 CTTATGAAGATGCTCAATCTGAATTTATTTGGCTACCTAAAGAAATGCAATATTTAGT 3831
 Oy 3632 ttccatttgcatagtgtgtgtgtctatagacaatatttaaatgaaatgttctt 3691
 Db 3832 TTTCCATTTGCATGATGTTGTGTCTATAGATGATATTTAAATGAAGTTTGTTT 3891
 Oy 3632 aaatattttaaagtgaaagactgtttcagctcttattatgtatagactttat 3751
 Db 3892 AAATATTTTAAACATGAAGACTGTTTACAGCTCTTTTAATGTACATAGCTTTTAT 3951
 Oy 3752 gtaat-----ctggcataattttgtgagacctttaaagctgataattcttccaaact 3808
 Db 3952 GTAAATTAAGTGTCTCAATCTGAATTTATTTGGCTACCTAAAGAAATGCAATATTTAGT 4010
 Oy 3809 ttgaatacaaaacagctgttataact 3836
 Db 4011 TTGAATATCAAAACAGGTTTTTACT 4038

RESULT 3

AL133551

LOCUS Human DNA sequence from clone RP11-57G10 on chromosome 10 contains a J-domain containing protein (JDP1) isoform B, the SIRT1 gene (Sirt1-like proteins (sirutinins) type 1), part of a novel gene similar to KIA0032, two ribosomal pseudogenes, 2 CpG islands, ESTs, STSs and GSSs, complete sequence.

ACCESSION AL133551
 VERSION AL133551.13 GI:9407715
 KEYWORDS HMG; J-domain; RP12; RP121; SIRT1.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Wilson, S.
 TITLE Direct Submission
 JOURNAL Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT On Jul 23, 2000 this sequence replaced gi:9367364.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information

on the WormPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>

RP11-57610 is from the library RPC1-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>
VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-57610 The true
left end of clone RP11-474D14 is at 172652 in this sequence.

FEATURES

source	Location/Qualifiers
repeat_region	1..175940
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="10"
repeat_region	/clone="RP11-57610"
repeat_region	/clone.lib="RPC1-11.1"
repeat_region	1..2318
repeat_region	/note="L1PA10 repeat: matches 3815..6158 of consensus"
repeat_region	2338..2767
repeat_region	/note="L1MA4 repeat: matches 4946..5385 of consensus"
repeat_region	2768..3058
repeat_region	/note="AluJo repeat: matches 1..296 of consensus"
repeat_region	3059..3371
repeat_region	/note="L1MA4 repeat: matches 5385..5704 of consensus"
repeat_region	3372..3682
repeat_region	/note="AluSx repeat: matches 1..310 of consensus"
repeat_region	3683..4322
repeat_region	/note="AluSx repeat: matches 5704..6298 of consensus"
repeat_region	4346..4707
repeat_region	/note="L1MB5 repeat: matches 5798..6175 of consensus"
repeat_region	4872..4926
repeat_region	/note="L1MB5 repeat: matches 204..258 of consensus"
repeat_region	4880..4938
repeat_region	/note="L2 repeat: matches 2648..2731 of consensus"
repeat_region	5029..5146
repeat_region	/note="L2 repeat: matches 436..564 of consensus"
repeat_region	5150..5194
repeat_region	/note="MLT1F repeat: matches 2..50 of consensus"
repeat_region	5195..5754
repeat_region	/note="MLT2CB repeat: matches 2..50 of consensus"
repeat_region	5755..6125
repeat_region	/note="MLT2CB repeat: matches 1..453 of consensus"
repeat_region	6142..6207
repeat_region	/note="MLT2CB repeat: matches 42..454 of consensus"
repeat_region	7403..7494
repeat_region	/note="33 copies 2 mer ac 86% conserved"
repeat_region	7561..7922
repeat_region	/note="MLT1E repeat: matches 1..81 of consensus"
repeat_region	8283..8422
repeat_region	/note="MLT1E repeat: matches 180..568 of consensus"
repeat_region	8415..8907
repeat_region	/note="MER7A repeat: matches 191..336 of consensus"
repeat_region	8908..9107
repeat_region	/note="MER7A repeat: matches 11..527 of consensus"
repeat_region	9110..9171
repeat_region	/note="MER7A repeat: matches 1..211 of consensus"
repeat_region	9449..9750
repeat_region	/note="MIR repeat: matches 76..146 of consensus"
repeat_region	10489..10938
repeat_region	/note="AluY repeat: matches 1..304 of consensus"
repeat_region	11938..12442
repeat_region	/note="L1R2 repeat: matches 1..448 of consensus"
repeat_region	12362..12436
repeat_region	/note="CpG island"
repeat_region	/evidence="not experimental"
repeat_region	12947..13250
repeat_region	/note="tRNA-Ser-TCA repeat: matches 8..82 of consensus"
repeat_region	13888..13947
repeat_region	/note="AluY repeat: matches 1..303 of consensus"
repeat_region	14085..14387
repeat_region	/note="MLT1J repeat: matches 453..514 of consensus"
misc_feature	/note="AluSg repeat: matches 1..304 of consensus"
repeat_region	complement(14553..14948)
repeat_region	/note="match: GSS: Em:AQ113044"
repeat_region	14626..14764
repeat_region	/note="MLT1J repeat: matches 229..366 of consensus"
repeat_region	15268..15328
repeat_region	/note="MLT1J repeat: matches 113..172 of consensus"
repeat_region	15822..16120
repeat_region	/note="AluSx repeat: matches 1..306 of consensus"
repeat_region	16150..16463
repeat_region	/note="AluY repeat: matches 1..312 of consensus"
repeat_region	17598..18107
repeat_region	/note="L2 repeat: matches 1843..2390 of consensus"
repeat_region	18101..18225
repeat_region	/note="match: GSS: Em:AQ627100"
repeat_region	18230..18711
repeat_region	/note="match: GSS: Em:AQ627100"
repeat_region	19538..19830
repeat_region	/note="AluSg repeat: matches 1..294 of consensus"
repeat_region	20049..20256
repeat_region	/note="MER58A repeat: matches 12..224 of consensus"
repeat_region	20724..20880
repeat_region	/note="MIR repeat: matches 20..175 of consensus"
repeat_region	21560..21844
repeat_region	/note="AluSg repeat: matches 1..302 of consensus"
repeat_region	21917..22003
repeat_region	/note="L2 repeat: matches 2207..2295 of consensus"
repeat_region	21929..22382
repeat_region	/note="match: GSS: Em:AQ672646"
repeat_region	22135..22178
repeat_region	/note="22 copies 2 mer ac 81% conserved"
repeat_region	22288..22595
repeat_region	/note="AluSx repeat: matches 1..305 of consensus"
repeat_region	24457..24579
repeat_region	/note="L1R33 repeat: matches 396..521 of consensus"
repeat_region	24729..24815
repeat_region	/note="L1R33 repeat: matches 125..213 of consensus"
repeat_region	24992..25152
repeat_region	/note="FLAM_A repeat: matches 10..120 of consensus"
repeat_region	25153..25419
repeat_region	/note="AluJo repeat: matches 1..263 of consensus"
repeat_region	25420..25432
repeat_region	/note="FLAM_A repeat: matches 120..133 of consensus"
repeat_region	26638..26802
repeat_region	/note="MER5A repeat: matches 7..189 of consensus"
repeat_region	27004..27072
repeat_region	/note="Alu repeat: matches 238..306 of consensus"
repeat_region	27063..27175
repeat_region	/note="AluJo/FRM repeat: matches 144..250 of consensus"
repeat_region	27176..27447
repeat_region	/note="AluY repeat: matches 3..274 of consensus"
repeat_region	27453..27532
repeat_region	/note="L2 repeat: matches 2586..2705 of consensus"
repeat_region	28187..28431
repeat_region	/note="AluY repeat: matches 30..307 of consensus"
repeat_region	29113..29319
repeat_region	/note="L2 repeat: matches 2476..2693 of consensus"
repeat_region	29643..29942
repeat_region	/note="AluJo repeat: matches 1..293 of consensus"
repeat_region	30116..30180
repeat_region	/note="MLT1C repeat: matches 406..466 of consensus"
repeat_region	30181..30485
repeat_region	/note="AluY repeat: matches 1..305 of consensus"
repeat_region	30486..30861
repeat_region	/note="MLT1C repeat: matches 1..406 of consensus"
repeat_region	30889..31229
repeat_region	/note="L1MC5 repeat: matches 7547..7904 of consensus"
repeat_region	31681..32033
repeat_region	/note="L2 repeat: matches 2379..2709 of consensus"
repeat_region	32163..32593
repeat_region	/note="MER57A repeat: matches 1..434 of consensus"
repeat_region	32694..32976
repeat_region	/note="AluJo repeat: matches 1..298 of consensus"

repeat_region	33377..33567	/note="MER3 repeat: matches 3..201 of consensus"
repeat_region	33583..33734	/note="MER5A repeat: matches 9..189 of consensus"
repeat_region	33759..33812	/note="MER5A repeat: matches 128..181 of consensus"
repeat_region	33813..34125	/note="AluSq repeat: matches 1..296 of consensus"
repeat_region	34126..34234	/note="MER5A repeat: matches 21..128 of consensus"
repeat_region	34300..34598	/note="MLT1A1 repeat: matches 93..361 of consensus"
repeat_region	34899..35075	/note="LTR45 repeat: matches 349..525 of consensus"
repeat_region	35076..35177	/note="MLT1A1 repeat: matches 1..108 of consensus"
repeat_region	35183..35482	/note="AluSq repeat: matches 1..301 of consensus"
repeat_region	35484..35650	
Query Match	21.4%	Score 828.4; DB 37; Length 175940;
Best Local Similarity	73.2%	Prod. No. 5.8e-100;
Matches 1547; Conservative	0; Mismatches 371; Indels 196; Gaps 30;	

	Query Match	Similarity	21.4%	Score 828.4	DB 37	Length 175540
	Best Local	Similarity	73.2%	Pred. No. 5.0e-100		
	Matches 1547	Conservative	0	Mismatches 371	Indels 196	Gaps 3
QY 1904	aaagagcagatctgtaagcgccttgagggagcaatcaactgtctgtacaccaaactcg	1963				
Db 164087	AAAGTAACATTTTATTACTGATTTTTCAGGTAATCACTACTGTTTGTGGCACCMAATCG	164148				
QY 1964	ttacatattccacggctgcgtagatatactcaactctgaagatgaagtcctgtccctag	2023				
Db 164147	TTACATTTTCCATGGCGCTGAGGTATATTCAGACTCTGAAGATGACGCTTATCTCTAG	164206				
QY 2024	ttccgtgtggcagtaagatgagcaatggcaaatgccaagatgccaagtttaagaacccct	2083				
Db 164207	TTCTGTGTGGCAGTAACGTGATATGGACATGCGACAGTCCAGATTCAGATTAAAGAACCCAT	164266				
QY 2084	ggaagatgaaagtgaatcttgagaatctctcaaaacygctctgaaagtatatacggagagcc	2143				
Db 164287	GGAGAGTGAAGTGAATTTGAAGAATTTCTACAAATGCGTTTGAAGATAGCGTATGTCC	164328				
QY 2144	cgaatctgctggagagatcttgattctggagctgatatggaaggagatcaagagtcgttaatga	2203				
Db 164327	AGACAAGACCTGGAGAGAGCTGATTTGGGACGTAGTGAATGATGATCAAGAGCAATTAATGA	164386				
QY 2204	agccatgctcaagaagacgaagatggaagaatgtaactatccatccagcaaatcataca	2263				
Db 164387	AGCTATATCTGTGAACGGAAGTAAACAGACATGAATATCTATCCATCAAAACAAATCAT--A	164443				
QY 2264	ctatctgaagcctgcgcggaatctcagaagatctgctccacacgacatctggaaactctagatgca	2323				
Db 164444	GTTGTAATTAATGTGTCAGGTACAGGAATTTGTTCCACAGCAATTTAGGACTTACGATCTC-	164502				
QY 2324	aaaaatgaaatgcttactctgtgcaac---ttgaaacaagaatctgaagaatgatatat	2379				
Db 164503	-----AAAATGAAGTGTACTCTTGGAACCTCGAATAGACCAAGAAACAGAAAGTGAATATT	164560				
QY 2380	tatagacaggg--aaaatgatatg--cctctctggaatattcttaagtccaactctctgtc	2437				
Db 164561	TATAGGTTGGTAATAATGATTTGTTTTCATAGGATTAATTTTAACCTTATTTCTGTAC	164620				
QY 2438	ctgactctgtacatcaacaactctgtgtgact-----	2470				
Db 164621	TTGTACAACACCAACACGTAACCTTTTTTTTTTTTAAAAAAGTACGTAAGTATCTTC	164680				
QY 2471	-----tcaactctcttcaaggctcaattgtatgtatataactgtatca	2510				
Db 164681	AATCAGCTGTGTGCTCAAGACATCACTTTCTTTTAAAGTTCATTTGTATGAATTAATCATATA	164740				
QY 2511	tgatgtata-----atttgttttcttccaaagagtttcaactcttaagttt--ca	2564				
Db 164741	TGTGTATATATAAATTTTGTGTGTGTTGTTGCTAGAGAGTTTCAACATTTTAAAGCTTTTCAA	164800				
QY 2565	aaagccattggaatg-----ttaaagcaagggaacgcctctatctaaaccaaagaatgtc	2619				

Db	164801	AAAGCATTGGAAATTTAAATTAAAGTAAAGGAAGCAACCTATATGACCAAAATAGT	164806
QY	2620	attcaacacattctgtttgttaacaatgaaatagtttaagccctcaattctcgttcgt	2679
Db	164861	ATTT---TCACATTTTCTTTGGTAACATTTGAAGGTTTGGAAGTACCAAAATCTGTTAGCT	164917
QY	2680	gaactattatttttagacagtttaactcttttaaacacgtggaatttccaaacttgygc	2739
Db	164918	AAACTTTTGTATTCTTTAAACACAAATTATTTTAAACACGTGGCATTTTCCAAAC-TGtGGC	164978
QY	2740	agctaactttttaaatacagaatgacttgcatactgaa-----ggagtcagccagctg	2792
Db	164977	ACCTAACTTTTAAATCTCAAAATGACATGCGAGTGTAGTGAAGGAATCAACAAATATG	165036
QY	2793	tctggaagcactc-----aaacttgg	2813
Db	165037	TGGGAGAGACATCGGTTGTCTTTACTTTTAAAGTAATACTGTGCTTAAGAAATTTTCAG	165096
QY	2814	gtctgaatgtgtaagcgtactactgcaactgttttttgactgtctgcaagac--gtggtaa	2871
Db	165097	GATTATTGTATTGTACGTTCAAAATGAAGATGGCTTTTGTACTTCTGTGGACATGTACCA	165156
QY	2872	tgctcaacaagccctgcagactaaactgataaattgalt-----tgyaaatglttca	2925
Db	165157	TGTCTATATTGGCTCATTAACCTAACTAACTGAAGAAACAAATAATCTTTGGAAATGTTTCA	165218
QY	2926	gttgtcttagaagaatagttgctgtctatattagtcaccttagtttggaatattgcat	2985
Db	165217	GTGTGCTTTAAGAAACATTTAGTGTGCTGCTGGAT---CCCTTAGTTTGAATATTGTGCCA	165273
QY	2986	tgtttaattaaatacactaactgctgtagagccctgcaatagatcttccacaacaa-----a	3040
Db	165274	TGTGTTGTTTAAATACCTTAACCTGACACTGTGGTAGAGCTTGATGCTTTTCCACAAGTATTA	165333
QY	3041	tactgcacaagatgtaataatagcaaaagccttctgaaac--taataatgtaacttctactg	3098
Db	165334	AACTCCCAAAAGTAAATATGCCAAAGCCCTTCTCAATCTTAATAAATGAAGTACTTCACTG	165393
QY	3099	ggagagaagttaactatcttggaactgctgtcttccatctaagtaggaagaacatagagcctt	3158
Db	165394	GGGAAGTGTTAATATTTTGGACTGCTG-TTTTCCATTATAGAGAGAGACCAAGGCCCT	165453
QY	3159	taat--taaaagtcacaagatcatalaaga--taaatgttagctcaacagaagaatgtaactglt	3215
Db	165453	GATTATTAAGTCTCAAACTAATTAAGATGTAAATGTAAATTCAGCCCAAAAGTACATGTCT	165512
QY	3216	gactgttagagattggtgtgtaatgtatcccaagtglttagctgtgtatttgagaag--	3273
Db	165513	CCCATTTGGAGAGATTGTGTTAAATATACCAACGTGTAAGCCCTAGTATGATATAGGAGATGAAT	165572
QY	3274	-----aatacagatcccaatagccaatgaaactagttcctcag-----	3310
Db	165573	CATTGATGATGTAACTTGTAAATATAGCAGAAATAGTTAATGATGAATGAACCTAGTTCTTAATTT	165632
QY	3311	-----ttaattaaagcttagcttgcttaaaaactaaggatcaattcttccaaactgcaaga	3366
Db	165633	ATCTTTATTATTAAACCTTAGGCTGCTTTAAACATGAAGATCAACCTTTCTCAGCTGCAGAAA	165692
QY	3367	acctttagccttcaaacagttcaaacactgaa-----agtcagattttaaattac	3418
Db	165693	GCTTTAGTCTTTCAAGAAGTTCATACTTTATGAAATTTGCACAGTAAAGCATTTATTTTTTC	165752
QY	3419	agacttcttggaaactgcccacaatttaaatatt-----catgtyggatttagtaatt	3473
Db	165753	AGACCATTTTGGACATCATCTCTAAATTAATTAAGAATTCCTGTGTTGCTTTAGTATTT	165812
QY	3474	attacaac-----aaaatgatttgaatatagcgtgtctttatgtatgaataataccocglttaga	3532
Db	165813	ATTACAAATTAAGAGGTTTGAATATATGCTGTTCTTTATGCAATAAACACCCAGCTTAGGA	165872
QY	3533	caattacgcaagagagaaaagtattaaagta-gctcaattccactactaaagat----	3587

	Db	165873	CCATTACGCGCAGAGAAAAAATCGTATTGAATGGCCATTTCCTACTGTTATTAAGATGTCT	165932
OY	3588	--aacgtaatattatggctacatacgaagatgcagtatatattagttccattcgatg	3645	
Db	165933	CANCTGAAATTTATTTGGCTACACTAAAGAATGCAGTATATTAGTTTTCCATTTTGCAATG	165992	
OY	3646	atgtgttgtagcataggacaattttaattgaaatttggtttaatttatca	3705	
Db	165993	ATGTTTGTGTGCTATAGATGATATTTTAAATGAAAAGCTGTGTTTAAATATTTTTCACA	166052	
OY	3706	gtgaagactgttttcagctcttttatatttgacatagaactttagaat---	ctggca 3762	
Db	166053	GCGAAGACTGTTTTAGCTCTTTTATATGTGACATAAGCTTTTATGTATATTACTGGCA	166112	
OY	3763	tatgttgttagaccgtttaatgatgattatcttcctccaactttgaaatacaaaa	3822	
Db	166113	TATGTTTTGTAGACGTGTTAATGACTGGA-TATCTTCCTTCAACTTTGAATFACAAAAC	166171	
OY	3823	cagtgatttatact	3836	
Db	166172	CAGTGTGTTTTTACT	166185	
RESULT	4			
LOCUS	AP001595	154686 bp	DNA	PRI 29-APR-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 2L, clone:KB1622E1,			
ACCESSION	AP001595	region, complete sequence.		
VERSION	AP001595.1	GI:7670549		
KEYWORDS	HTG			
SOURCE	Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio			
ORGANISM	BAC library clone:KB1622E1.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 154686)			
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.			
TITLE	Homo sapiens genomic DNA, chromosome 2L, clone:KB1622E1,			
JOURNAL	APP-D1S292 region			
REFERENCE	Published Only in database (2000) In press			
AUTHORS	2 (bases 1 to 154686)			
TITLE	Shimizu,N., Kudoh,J. and Shibuya,K.			
JOURNAL	Direct Submission			
FEATURES	Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.			
	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular			
	Biology, 35 Shinanomachi, Shinjuku-Ku, Tokyo 160-8582, Japan			
	(E-mail:nshimizue@med.keio.ac.jp, Tel.:81-3-3351-2370,			
	Fax:81-3-3351-2370)			
	Location/Qualifiers			
source	1..154686			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/cell_line="FLEB14-14"			
	/cell_type="pre-pro-B cell"			
	/chromosome="21"			
	/clone="KB1622E1"			
	/clone_lib="Keio BAC library"			
BASE COUNT	47469 a 28509 c 30527 g 48101 t			
ORIGIN				
	Query Match 7.4%; Score 286; DB 48; Length 154686;			
	Best Local Similarity 80.6%; Pred. No. 5.6e-29;			
	Matches 429; Conservative 0; Mismatches 75; Indels 28; Gaps 7;			
OY	3332	cttaaactaggatcaatttctcaactgcgaaccttttagccttcaaacagttcac	3391	
Db	119381	CTTAAACTAGAGATCAACTTCTTAGTCGCAAAAGCTTCGACCTTTCAGAGATCAT	119440	
OY	3392	aacctagaagc-----agattattattaagactcttttggaaatgccccc	3443	
Db	119441	ACTTTATGAAATTTGCACATTAAGCATTTATTTTGAACCATATTTTGAACATTA	119500	

QY	3444	aactaaatatactcagtg-----ggtttgattatcttacc-aaaaatgatttgaata	3497
Db	119501	AATTAAATAAAGTATTTGTAAGTACTGCTTTAGTATTTGATTAACAGTAAAGGGTTTAAATA	119560
QY	3498	tactgtcttcttaagcctaaataaccgaactgaaccattactgcgcagagagaaaagta	3557
Db	119561	TAGCGTCTCTTATGCAATAAAATACCAGTAGACACTACTGCCAGAGAAACATCG	119620
QY	3558	ttaaagta-gctcattccctaccctaaagat-----aacgcaattcatttgcctacact	3610
Db	119621	TATTGAATGGCCATTTTCCCTACCTCAAAAGATGTCGAATCTGAATTTATTTGGCTACACT	119680
QY	3611	aagaagatgcagatattatgaattttccatttgcgatgtgtgtgtgtgtgtatagaacaact	3670
Db	119661	AAAGAAATGCACTAATATTTTACTGTTTCCATTTGGCATTAATGTTGGTGTGCTATTAATATTAAT	119740
QY	3671	ttaaattgaanaaatltgttttaataatttttccagtagaagactgttttcaactcttct	3730
Db	119741	TTTAATTTGAAGAAGTTGTGTTTTTAATATTAATTTTTCATATGAAGACTGTGTTTCACACTCTTTT	119800
QY	3731	ataattgaatagactctt-----atgtaatctgtgcataagtgtttgtagacggttcat	3784
Db	119801	ATATTGTGCACATAGTTTGTATATGTAATCTACTGCAATGTGTTTTTGTGTGACTGTTTAAT	119860
QY	3785	gactgagattactctccctcccaacttttgaatatacaaaaacagtggttttact	3836
Db	119861	GACTGGA-TATCTTCCTCCATATTTTGAATACAAACCGTGTATTTTACT	119911
RESULT	5		
LOCUS	AP001696	AP001696 340000 bp DNA PRI	30-MAY-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21q, section 40/105.		
ACCESSION	AP001696	AL163241	BA000005
VERSION	AP001696.1	GI:7768673	
KEYWORDS			
SOURCE	Homo sapiens DNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
	Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Tokoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Patterson,D., Reichwald,K., Rump,A., Schlichthabel,M., and Schudy,A.		
TITLE	The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium		
JOURNAL	Nature 405 (6784), 311-319 (2000)		
MEDLINE	2289799		
REFERENCE	2 (bases 1 to 340000)		
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Matanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Tokoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schlichthabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Barand,P., Scherfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Henrich,S., Riesenmann,L., Dagnand,E., Wehrmayer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and Yasp,M.L.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-Apr-2000) to the DDBJ/EMBL/Genbank databases. The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Kelo University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)		

COMMENT

On May 30, 2000 this sequence version replaced gi:7717294.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan.
* e.mail: hattori@gscc.riken.go.jp/
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gsc1-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
160-8582, Japan,
* e.mail: nshimizue@mb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163241: Submitted (10-Apr-2000).

FEATURES

source

Location/Qualifiers

1..340000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

<1..19901

/note="Accession No. AL109967"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q21.2"

/clone="Q22F1, 5' partial"

/clone.lib="L21NCO2-Q Cosmid library"

<1..38292

/note="Accession No. AP000230"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q21.2"

/clone="pR4F3, 5' partial"

/clone.lib="CMP21 Pl library"

<1..108797

/note="Accession No. AP001595"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q21.2"

/clone="KB1622E1, 5' partial"

/clone.lib="Keio BAC library"

1965..4225

/note="L1M1"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

complement(4226..4526)

/note="AluJb"

/rpt_family="SINE/Alu"

/rpt_type="DISPERSED

4527..4970

/note="L1M1"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

4965..6293

/note="L1MA7"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

complement(8517..8601)

repeat_region

/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED
8627..8927
/note="AluSc"
/rpt_family="SINE/Alu"

repeat_region

/rpt_type="DISPERSED
9686..9706
/note="(A)n"
/rpt_family="Simple_repeat"

repeat_region

/rpt_type="TANDEM
10415..10509
/note="MIR"
/rpt_family="SINE/MIR"

repeat_region

/rpt_type="DISPERSED
complement(10578..10703)
/note="Charlie2"
/rpt_family="DNA/MER1_type"

repeat_region

/rpt_type="DISPERSED
complement(10717..10903)
/note="Charlie2"
/rpt_family="DNA/MER1_type"

repeat_region

/rpt_type="DISPERSED
complement(11074..11361)
/note="Charlie2"
/rpt_family="DNA/MER1_type"

repeat_region

/rpt_type="DISPERSED
11690..11834
/note="MIR"
/rpt_family="SINE/MIR"

repeat_region

/rpt_type="DISPERSED
11896..11995
/note="LTR33"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
complement(12160..12314)
/note="HERV16"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
complement(12411..12703)
/note="HERV16"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
complement(12698..13057)
/note="HERV16"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
complement(13061..13130)
/note="LTR16A1"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
13131..13436
/note="AluY"
/rpt_family="SINE/Alu"

repeat_region

/rpt_type="DISPERSED
complement(13437..13810)
/note="LTR16A1"
/rpt_family="LTR/Retroviral"

repeat_region

/rpt_type="DISPERSED
14211..14514
/note="AluY"
/rpt_family="SINE/Alu"

repeat_region

/rpt_type="DISPERSED
complement(15748..16068)
/note="AluY"
/rpt_family="SINE/Alu"

repeat_region

/rpt_type="DISPERSED
complement(16549..16848)
/note="AluY"
/rpt_family="SINE/Alu"

repeat_region

/rpt_type="DISPERSED
17410..17494
/note="MER20"


```

/rpl_family="DNA/MERL_type"
/rpl_type=DISPERSED
complement(17546..17810)
/Note="MULTI1"
/rpl_family="LRR/MALR"
/rpl_type=DISPERSED
18069..18215
/Note="(TATC)n"
/rpl_family="Simple_repeat"
/rpl_type=IANDEN
complement(18645..18970)
/Note="AluJb"
/rpl_family="SINE/Alu"

Query Match      7.4%  Score 286; DB 48; Length 340000;
Best Local Similarity 80.6%  Pred. No. 4.8e-29;
Matches 429; Conservative 0; Mismatches 75; Indels 28; Gaps 7;

QY 3332 ctaaacactagagatcaattctcacaagcagaactttaagccttcaaacagttcac 3391
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73492 CTAAACAGTAGACATCACTCTTCTAGCTGCAAAAGCTTACCTTTCAAGAAGTTGAT 73551

QY 3392 accctcagaagtc-----agattatattacagactcttggagacatgcccaca 3443
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73552 ACTTATGAAATGACACATTAACCATTTATTTTCAACCATTTTGAACATTAATCTCTA 73611

QY 3444 aatttaaatattcatgty-----ggttggattattataa-aaaaatgattgaata 3497
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73612 AATTAATTAAGATWTCGATGACTGCTTACTATTTGATTAAGTAAAGGGTTGAATA 73671

QY 3498 taagcttctttagatgaataaacacagcttagacattacgcaagagagaagaagta 3557
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73672 TAGCTTTCTTTATGCAATAAATACCAAGCTAGGACAGTACTGCGAAGAAAACATCG 73731

QY 3558 ttaagta-gtccattccctaccctaaagaat-----aactgaattattctggctacac 3610
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73732 TATTGATAGGCGATTTCCGACTAAAGATGTCGATCTGAATTTATTTGGCTACACT 73791

QY 3611 aaagaatgcagatatttagtttccatctgcatgcatggttgcctacatacaaatat 3670
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73792 AAGAAATGACGATWATTTAGTTTTCATTTGCAATGATGTTGTCTATAGATATAT 73851

QY 3671 ttaactggaataatgttttaaatatttttcaagtgaaagactgttttcaagctcttt 3730
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73852 TTAATTTGAAAATTGTTTAAATTTATTTTTCATTAAGAGACTGTTTTCAGCTCTTTT 73911

QY 3731 atatgtacatagactttt-----atgtaatctggacatagtttttagacagcttaat 3784
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73912 ATATGTGCATAGTCTTTTATGTAATCTGATGATGATGTTTGTAGACGTTTAA 73971

QY 3785 gaatgaatatctctcctcaactttggaatacaaaaaacagtgctttatact 3836
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73972 GACTGGA-TATCTTCTCCCAATTTTGAAGTACAAAACAGTATATTTACT 74022

RESULT 6
AC022592/c 86516 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-22D9, LOW-PASS SEQUENCE SAMPLING.
AC022592
AC022592.2 GI:9124033
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86516)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens, clone RP11-22D9
JOURNAL
Unpublished
2 (bases 1 to 86516)
REFERENCE
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beka,F.,

```

TITLE JOURNAL

COMMENT

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choeopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Lien,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.-J.,
Zimmer,A., and Zody,M.

Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910733.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4104
Center clone name: 22_D_9

* NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 854: contig of 854 bp in length
* 855 954: gap of 100 bp
* 955 1827: contig of 873 bp in length
* 1828 1927: gap of 100 bp
* 1928 2795: contig of 868 bp in length
* 2796 2895: gap of 100 bp
* 2896 3760: contig of 865 bp in length
* 3761 3860: gap of 100 bp
* 3861 4723: contig of 863 bp in length
* 4724 4823: gap of 100 bp
* 4824 5679: contig of 856 bp in length
* 5680 5779: gap of 100 bp
* 5780 6625: contig of 846 bp in length
* 6626 6725: gap of 100 bp
* 6726 7575: contig of 850 bp in length
* 7576 7675: gap of 100 bp
* 7676 8524: contig of 849 bp in length
* 8525 8624: gap of 100 bp
* 8625 9475: contig of 851 bp in length
* 9476 9575: gap of 100 bp
* 9576 10437: contig of 862 bp in length
* 10438 10537: gap of 100 bp
* 10538 11366: contig of 829 bp in length
* 11367 11466: gap of 100 bp
* 11467 12291: contig of 825 bp in length
* 12292 12391: gap of 100 bp
* 12392 13257: contig of 866 bp in length
* 13258 13357: gap of 100 bp
* 13358 14229: contig of 872 bp in length
* 14230 14329: gap of 100 bp
* 14330 15183: contig of 854 bp in length

FEATURES
source
1. 68476
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="34A-34E"
/clone="BACR32012 (D1025) RPCL-98 32.0.12"
/clone_lib="RPCL-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBAC3.6"
BASE COUNT 18340 a 15210 c 15061 g 17814 t 2051 others
ORIGIN

58781 58860: gap of unknown length
* 58861 67308: contig of 8448 bp in length
* 67309 67388: gap of unknown length
* 67389 67887: contig of 499 bp in length
* 67888 67967: gap of unknown length
* 67968 68476: contig of 509 bp in length.
Location/Qualifiers

QY 680 gctgtgcaatgttataatataatccttcagaaaccccaagcggaagaaagaaagaaag 739
|||||
Db 22495 GCTGTGGGACTATTTGGCCCACTGTTGAACAGACCGCGGCAACCAAGCTGCCCTC 22436
QY 740 tatcaataacatctgaagatctgtgaagctacgtgagagtgtaaaagataatagtctt 799
|||||
Db 22435 AGTGAACACCTTGACACATGTCTATCATCTTTGGTCAACAAATCAGAAAGATCATGTGCT 22376
QY 800 gactgagctgtgggttctgtctcgtgtgagatcctgactcagatcaagagagagat 859
|||||
Db 22375 AACGGGAGCGGAGATCCGCTCTCCGCAATCCGCAATCCGCTCGGTCACCAATGGCAT 22316
QY 860 ctgtgcgcgtcgtggtgagatcctccagaccccaagcctcaagcctatggtatata 919
|||||
Db 22315 ATATGCCGATGGCCCATATTTTCCCGATCTGCCGATCCGAGCCGACCAATGTTGATAT 22256
QY 920 tggatatttagaaagaaagcccaagcattcctcaagtttgcaagaaagataatcccg 979
|||||
Db 22255 CAACCTACTTCAAGAGGATCCAGACGCTTCTAACAAGTTGCCCGCAGATATATCCCGG 22196
QY 980 acagttcagcgcgtctcgtgtgcaaaatcataagcttgcagataaggaagaaact 1039
|||||
Db 22195 CGAGTTTCACCCCTCACACCCCTGCATCGTTTCATCAAAATGCTGGAGACCAAGGCAACT 22136
QY 1040 acttcgaatataatataatataatagatccttcagagcaggttcgaaagatccaaagat 1099
|||||
Db 22135 GTTGGCGCAACTACACAGAACATCGACACCTTCGACGGGTGGACGATTCAGCGAGT 22076
QY 1100 ccttcagtgctcagtgcttccttcgacacagacatcttcgctgattgtaatacaagtctga 1159
|||||
Db 22075 AATCGAATGTACAGGCTCTTTTCAACGCGCTCTGCACCAAGTGTGTTTCAAGTGCAG 22016
QY 1160 ttgtgaagctgttcgtgtgagagacatttaacagagtagtctcgtgcctcaggtgcc 1219
|||||
Db 22015 CGGTGACGCCCTGGCGGCGGACATATTTGCCACGAAATTCGGTGTGCCGCGAGTGCA 21956
QY 1220 agctgtatgag 1229
|||||
Db 21955 GCCCAATTAAG 21946

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

HTG.
fruit fly.
Drosophila melanogaster
Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
Pterygota: Neoptera: Endopterygota: Diptera: Brachyera:
Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
1 (bases 1 to 261612)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,V., Chame,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Basu,A.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriere,S.,
Fleischmann,M., Fowler,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Idegam,C., Jalali,M., Kalush,F., Karpis,G.H., Ke,Z.,
Kensiston,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kraavitz,S., Kulp,D., Lai,Z., Lascko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mittel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merklou,G., Milshina,N.V., Mobarry,T.C.,
Morris,T., Moshier,A., Mount,S.M., Moy,M., Murphy,B., Moberly,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Paclet,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Sytkas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,R.N., Zhong,H., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M., and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 261612)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7298000.
Location/Qualifiers
1. 261612
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
join(<11374, 11418, 11877, 12384, 12747, 13008, 13067, 13881,
16119, >16174)
/gene="CG5122"
/product="CT16417"
/db_xref="FLYBASE:FBan0005122"
/db_xref="FLYBASE:FBgn0032471"
/db_xref="FLYBASE:FBgn0032471"
/gene="CG5122"
/map="34A7-34A8"
/db_xref="FLYBASE:FBan0005122"
/db_xref="FLYBASE:FBgn0032471"
join(<11374, 11418, 11877, 12384, 12747, 13008, 13067, 13881,
16119, 16174)
/gene="CG5122"

COMMENT
FEATURES
source
gene
CDS

RESULT 10
AE003639
LOCUS AE003639 261612 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386055 section 32
of 63, complete sequence.
ACCESSION AE003639 AE002690
VERSION AE003639.2 GI:10728758

/note="CG5122 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0005122"
 /db_xref="FLYBASE:FBgn00032471"
 /protein_id="AAF53242.1"
 /db_xref="GI:7298001"
 /translation="MGCGDHPKKKKKKYLAPQVPRKCIKSEIRILADANDPNPKPEL
 KFPALPLQSTASTLSTIEPLINEELKEDQAKKRFVNDGAKLQDLLEEEEGEDTD
 NMVTPMTTSALITQVAPLTVFSSGSELPQIEFKDINDFLNTPAKIAYGCEKEKQVLV
 DONKIPVMOGKHOLDNSOFKIPEDTVTEIQSLFVYVCLDNCVPTPGSEEEKIVADOLH
 LYGCEGROASARKMDTKTQILVNPNGSLAGFCYEHSPPADCOPLAMLDVFOYLAKEPNE
 LGECEGARKKYTSNLLKFPQNECINMLGCAARNIQRITVSOLOMVNINECYGKCFP
 IKAQGLNDPSYIQMALSLATYSMYKITEAQVESALIRIEHSGRETTIRSTNSEKTTVL
 LMQSKRABSVETKLAQLQAVDAHEKLIKRAINLGGQIDRHLFGLQEMALEGMDTIPPEK
 FPSKGRVTSERFQLTQVATSNEGFMAYGPLLSDGYGVCCYNPESKITYFVISMKSC
 KEMITREFKAKIKSLINDMRKLIILLTGGERAGEHRCNCQAYSPATWMSMP"
 complement(21672..22079)
 /gene="CG9928"
 /product="CT72834"
 /db_xref="FLYBASE:FBan0009928"
 /db_xref="FLYBASE:FBgn00032472"
 complement(<21672..22079)
 /gene="CG9928"
 /map="34A7-34A8"
 /db_xref="FLYBASE:FBan0009928"
 /db_xref="FLYBASE:FBgn00032472"
 complement(21732..22052)
 /gene="CG9928"
 /note="CG9928 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0009928"
 /db_xref="FLYBASE:FBgn00032472"
 /protein_id="AAF53243.1"
 /db_xref="GI:7298002"
 /translation="MQSTIRLLVYGLIYTAIGASGFPRONINLGEGRTPSSGGINL
 WSDSVHSSDDDTLKLLESLDNSGDRNMKNDAVRNSGFTDTSAGDSKEDANGFOF
 HILG"
 complement(<22494..222784)
 /gene="CG16978"
 /product="CT35582"
 /db_xref="FLYBASE:FBan0016978"
 /db_xref="FLYBASE:FBgn00040972"
 /evidence=not_experimental
 complement(<22494..222784)
 /gene="CG16978"
 /map="34A7-34A8"
 /db_xref="FLYBASE:FBan0016978"
 /db_xref="FLYBASE:FBgn00040972"
 /evidence=not_experimental
 complement(22494..22784)
 /gene="CG16978"
 /note="CG16978 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0016978"
 /db_xref="FLYBASE:FBgn00040972"
 /evidence=not_experimental
 /protein_id="AAF53244.1"
 /db_xref="GI:7298003"
 /translation="MMSRIILLIFGTICALLTNGPHPLIEDPEGTNDALSSGSG
 YNSAEFNNAMDELIRNQENDAHENSTDPVSLDDDETTLDPMDDITTEPAT"
 join(37997..38069,38131..39305)
 /gene="PK34A"
 /product="Cn16567"
 /db_xref="FLYBASE:FBan0005182"
 /db_xref="FLYBASE:FBgn00028410"
 <37997..39305
 /gene="PK34A"
 /note="CG5182"
 /map="34A8-34A8"
 /db_xref="FLYBASE:FBan0005182"
 /db_xref="FLYBASE:FBgn00028410"
 join(38066..38069,38131..39305)
 /gene="PK34A"

```

/note="PK34 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005182"
/db_xref="FLYBASE:FBgn0028410"
/protein_id="AAF53245.1"
/db_xref="GI:7298004"
/translation="MEKKEKOKPLKRAHQVPLIIPKNSVISTYAVGLCEPALVRI
EKVDLIGSGSEFRVYQAHVNESEETVAQKQLYNKLSQGEAEIMGQLKHNNIVRLI
MHSYSIDPVDYVILVMEYWPMTLLDIYNHVLVQLPAEALINVRILSYOMEGFRLI
YHLHGISHRDVYPENLLIDNOKMYKLDSFESALVLOPBEISYICSRILYRAEELF
AGYELSCAVDIMSAGVLALELLKGYEPFSSKHROKRLIIVNMLGTDGLERAEIL
SKCGSLHPRTRPSMYNLINAVPODLCGLNSCFIYBAARISYPMACHSGSIDEL
RMDMALPMENGNLPLPLEDFNSLEKOTDPLWNLPLIHLSHEDTISVEAEAV
"
join(39714..39773,39980..41549,41607..>42293)
/gene="CG5204"
/product="CPI6653"
/db_xref="FLYBASE:FBan0005204"
/db_xref="FLYBASE:FBgn0032473"
<39714..>42293
/gene="CG5204"
/map="3A8-34A8"
/db_xref="FLYBASE:FBan0005204"
/db_xref="FLYBASE:FBgn0032473"
join(39993..41549,41607..42293)
/gene="CG5204"
/note="CG5204 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005204"
/db_xref="FLYBASE:FBgn0032473"
/protein_id="AAF53246.1"
/db_xref="GI:7298005"
/translation="MELPSWERSGDLRVLRISLVGAPLQOSIEGAGAVLPMNSOV
OPLIGDFLEOLLEQYKANPMEFPLSMGFPASAPAGDLAKEMENSLKYNHPCN
NNDDLPCKIRBNSCNOCAEFQNPSPHLSHLEDVNGEPIVINECLCLYASHFO
KLVRRMKHGGSDGDAVQSGAQPRKRNILREVRKRLIEESTEGQAGTQCDLS
VLRMLONGPRLQLVELDQOEMQLASVQAVNRQOEMLOLQIVESHDNLFPMAYEL
OTKLMPKPAESILKEQNSDSEETPYKMLDRTKHNHAGHFKSCDCEFDIKSLT
REKTKYHSMPLIKSCSDPHTPEKMLDRTKHNHAGHFKSCDCEFDIKSLT
IHESHVHPMPYHOMGNSRDEALDVOOGSGSKPETFKNGATVASTSLPRTS
GIVCHCCKRVGNAMQNLNHLQVCTLALHNTTQOASINAEVLDHDEFPNAPDLS
GVEFAPIGTEYVLEPEEDLAPLKVFCQPHICSTFAATSRHVIIVGHUNKRP
ECSLAYRSMRMWDITKILKALDRSHNOAVLMDGTGRNAYKANQVLTMMKVS
AEOLNDSGMRGTGEMIVMPPEKLDHNMETEEIIEVDSAHSTSLDLRPRDOTE
DIAGNSDELPOGATKEPNLEPLSPKSNSSQVMPKPKGSPWNLKKTGQGSDETTSA
DESNSD"
complement(join(42737..43390,43503..43949,44012..>44163))
/gene="CG9688"
/product="CT27734"

Query Match      5.3%: Score 20.4; DB 31; Length 26162;
Best Local Similarity 60.7%: Pred. No. 2.4e-18;
Matches 334; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

mRNA

680 gctgtgacagattgtatataataccttcaaaacccaacgaacggaagaaagaaga 739
      |||||
45738 GCGTGTGGACTATTGGCCACCTGTTGAAGCGGAGCGGACCAACAGCTGGCTTC 45797
      |||||

740 tatcaataaatgtaagatgctgtgaagttaactgcaagagtgtaaaagataatagtttc 799
      |||||

45798 AGTGAACACCTTCGACGAGTGTATCGTTGGTCAAGAAATACACAAMAATATTTCTGCT 45857
      |||||

800 gactggaagctggagttctgtctcctgttggaattcttgaactcaagaagacggtat 859
      |||||

45858 AACGGAGCGCGGAGTATCCGCTCCGCGGCAATCCGGACTTCGCTCCACCAATGGCAAT 45917
      |||||

860 ctatgcctgccttcggtgagcttcccaaacctcccaagaccctcaagccatgttgatat 919
      |||||

45918 AATATGGCGCATTTGGCCACGATTTTCCGATCTGGCCGATTCGCGAGGCCATGTTTGATAT 45977
      |||||

920 tgagtaatttagaaagaccacacatcttcaagtttgcaagaagaaatatatccgg 979
      |||||

45978 CAACCTACTTCAAGAGGATCCAGCAACCGTTTACAAAGTTTGGCCGCGAGATATATCCCGG 46037
      |||||

```

OY	980	aacatgcaccgctcgtctgtccacaactttagcttgcttcacagtaagaagaaact	1039
D6	46038	CGAGTTTACGCCCTCACCTTGCAATCGTTTCATCAAATGCTGGAGACCAGAAGAACT	46097
OY	1040	acttcgaattataccaatatagataaccttggagcagglttgcaaggaatccaaggat	1099
D6	46098	GTTGGCACAATACACAGAACATGCAGACACCTCGAGGGGNGAGGCAATTCACGAGT	46157
OY	1100	ccttcagtcatcgcttccctttgcaacgcatcttcgctgatattgtaatacaagtga	1159
		-	
D6	46158	AATCAGAGTGTCACGGCTCTCTTTCAACGCCCTCGACACCAAGTGTCTTTCAATGCNA	46217
OY	1160	tgtgaagctgtctgtggagacattttaatcaagttagtccgtgcccagggtgcc	1219
D6	46218	CGCTACGCCCTCGCGGGGACATATTTCGCCAGCGAATTCGCGTCCGCCAGTGCCA	46277
OY	1220	agctgatgag	1229
		-	
D6	46278	GCCCCAATAAG	46287
RESULT	11		
CERILIA8/C	LOCUS		
DEFINITION	CERILIA8 27332 bp DNA INV 13-JUL-2000		
ACCESSION	Caenorhabditis elegans cosmid RLIA8, complete sequence.		
VERSION	270310		
KEYWORDS	HTE; Ankyrin; Isoleucyl-tRNA synthetase;		
SOURCE	UDP-glucuronosyltransferase; Yeast regulatory protein SIR2 like.		
ORGANISM	Caenorhabditis elegans.		
	Caenorhabditis elegans.		
	Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabdilitida;		
	Rhabdilitoidae; Rhabdilitidae; Pelodertinae; Caenorhabditis.		
	1 (bases 1 to 27332)		
REFERENCE	no one.		
AUTHORS	Genome sequence of the nematode C. elegans: a platform for		
TITLE	investigating biology. The C. elegans Sequencing Consortium		
JOURNAL	Science 282 (5396), 2012-2018 (1998)		
MEDLINE	99069613		
REMARK	The C.elegans Sequencing Consortium.		
	Erratum:[[published errata appear in Science 1999 Jan		
	1;.283(5398).:35 and 1999 Mar 26;283(5410):2103 and 1999 Sep		
	3;.285(5433):.1493]]]		
	2 (bases 1 to 27332)		
REFERENCE	Cummings,P.N.		
AUTHORS	Direct Submission		
TITLE	Submitted (30-MAR-1996) Nematode Sequencing Project, Sanger Centre,		
JOURNAL	Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,		
	Washington University, St. Louis, MO 63110, USA. E-mail:		
	jess@sanger.ac.uk or rwenematode.wustl.edu		
	Coding sequences below are predicted from computer analysis, using		
	predictions from GeneFinder (P. Green, U. Washington), and other		
	available information.		
COMMENT	For a graphical representation of this sequence and its analysis		
	see:-		
	http://webpage.sanger.ac.uk/cgi-		
	bin/displayrbdb-wormaceclass-sequence object-RLIA8		
	Current sequence finishing criteria for the C. elegans genome		
	sequencing consortium are that all bases are either sequenced		
	unambiguously on both strands, or on a single strand with both a		
	dye primer and dye terminator reaction, from distinct subclones.		
	Exceptions are indicated by an explicit note.		
	IMPORTANT: This sequence is NOT necessarily the entire insert of		
	the specified clone. It may be shorter because we only sequence		
	overlapping sections once, or longer because we arrange for a small		
	overlap between neighbouring submissions.		
	IMPORTANT: This sequence is not the entire insert of clone RLIA8.		
	It may be shorter because we only sequence overlapping sections		
	once, or longer because we arrange for a small overlap between		
	neighbouring submissions.		
	The true left end of clone RLIA8 is at 1 in this sequence. The true		
	right end of clone RLIA8 is at 11438 in		

sequence 270686.
The true left end of clone R10H10 is at 27227 in this sequence. The start of this sequence (1..110) overlaps with the end of sequence A1023841.
The end of this sequence (27227..27332) overlaps with the start of sequence 270686.
Location/Offshifters

1..27332
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="R11A8"
complement(join(894..1251,1301..1470),1524..1682,1733..2503,2683..2800,2851..2933))
/gene="R11A8.1"
complement(join(894..1251,1301..1470,1524..1682,1733..2503,2683..2800,2851..2933))
/gene="R11A8.1"
/codon_start=1
/protein_id="CAA94366.1"
/db_xref="GI:3879117"
/db_xref="SPTREMBL:Q21923"
/translation="MSEKGFNYPKRENQLQINAQVESSQTSRGVTFEYDRPSYNGKITYTSAPKKKRRPLFRFRSPAPKEVAKRLNRIDRSRDVLAWSIDSGPTNHPDNDDELICTGKYODESNENNSVETIENSNCYEETIERENREDDCLPGSSNSHLPNDHVENNYLPTANSFHESEKFPDDTADSDCNFTDDSFLEQVGHSEVETSYGGYODEDTIESTCTFTETAKDFOAMKHQTESKQENEDTIFEDMDSDNOKESELLDEVENCATANLEKRLDVSIDELNMLSEQKDRDSRSVLYLEKRYCENADLMILKRLRDVETKSTDFLLPLQAGNDSTIEVEYIEKRYEVCQIPLAENNTRENRPTEPFDNADKDFEERYENEMKTEPLSNPELFTDRLNKKRIEKLADVAEMQIRRIIEEFNRNPLTAPFAFPFYQNNFQMAPONVGEPSLNYPVOANRPNPFRVYONITFFNLVNSTYFPVNPFLNPPEPPFNPNLVDSLNGSTFNPINR"
join(3917..4034,4086..4179,4327..4970,5269..5580,5691..5911)
/gene="R11A8.2"
join(3917..4034,4086..4179,4327..4970,5269..5580,5691..5911)
/gene="R11A8.2"
/note="cDNA EST EMBL:M75779 comes from this gene"
/codon_start=1
/protein_id="CAA94366.1"
/db_xref="GI:3879118"
/db_xref="SPTREMBL:Q21924"
/translation="MVEQTTSEPAAGSMPIKISFGVKRREKNVQAPAKAVIEVDLSDDEEENKEMERAKARRKRVHMEEDGTVOGDIDKPKEAIVIPWVYEDHWRTQKLIKEKAKALGLTEERAKALTLVPLNGIDGENEESGEGKIVGEDRGDTEADYSAPJIESFGLAILRCCMKDDGGIGKNPQKVALKLPNRPGDGLKPNKPVGCKNKNKMGESKAEERKLEEKVSGFTKYVDGRNKKGYAKYEGKDDDSNLSLFTITATIGKTMKQIYAAYAKSSAEYEDSKCLINKSEYDKERDRLETERKKLESPSTSTQSSKDKSKSSSSKHDKNSSKYEENDKMVATGDLVRFIDEDFERGSLYEDKVIIVAGGNDVITIEDDRDVTATYFEDVQSWLEIYPREIGELMIVAGKRSQALVMDKDKREKVTARVATVNDVYTAFFEDVCSVIRHEEDEY"

gene
join(6571..6937,7284..7368,7419..7534,7586..7860,8084..8535,8862..9087)
/gene="R11A8.3"
join(6571..6937,7284..7368,7419..7534,7586..7860,8084..8535,8862..9087)
/gene="R11A8.3"
/note="predicted using GeneFinder
Similarity to Human UDP-glucuronosyltransferase (SW:UDLF_HUMAN)"
/codon_start=1
/protein_id="CAA94365.1"
/db_xref="GI:3879116"
/db_xref="SPTREMBL:Q21922"
/translation="MTQFLFKLLVCTTINEEYTLNLTFVPTLSHSHISFNKLAOLLATSGHOVTVLAQVDALLIENATDEFAVLAKKGVGRGHLROYLWSPGCEYEDSSP
NPRLFYLLKVSCTGFYACONMMADHFLSLQOQKDVGLVEQYDSCGFIKFSIGIG
ENTWLAATAIYRQOAEALGVHLPYSVLEPAFHSDMKSPFOKVINILIQATSPVLI
DVFOGOSRIFRQFEGDIGAISQETSSVILNSVPEFPMSPVSHQSPNIGITDKRR
KLDIPYQSIADSAENGFVLYSRGILARTYDMTSMQRIPTDSAPRSHITFYKYEI
SSNSTVSDNVILTPMIPOLPLAAHNNKTIITHGMSSTLETTMMSKPMILPFLP
DHAANSVAESKGAVALIDLMKRLRSRVVAHNTIITLNPKTQCKEFSRNFSDTPH

CDS

BASE COUNT

10680 a

6063 c

5794 g

10941 t

Query Match	3.5%	Score 135.4	DB 32	Length 33478
Best Local Similarity	55.7%	Pred. No. 3.8e-09		
Matches 317	Conservative 0	Mismatches 226	Indels 26	Gaps 2
QY 964	aggaatatatccggagcaagcttcacgcgtctctgtcacaataatcatagcttgcag	1023		
Db 2377	AGGAATATCTTCCAGCAGAGATTCGCACCATCTGTTTCAATCGATTTATCAAGAGCTTG	2436		
QY 1024	ataagaagaanaactacttcgaatttactccaatatatgatcccttgagcgggttg	1083		
Db 2437	AAACATCTGTGTCGTCTTCTTCCAAACTCACTACCTCAAAATATCGATTCATTTGGATCAGACTG	2496		
QY 1084	cagg-----aatccaagaatcccttcagcttcagcttcgtcttcgacacagacatct	1133		
Db 2497	TAGCATTCAGACTGTATTTAAAGAGTTGTAGAAATGTCAATGATCCTTTTCAAAATGTACA	2556		
QY 1134	tgcctgatttgtaataacaaagtatgtctgtaagctgttcgttgagacattttaacag	1193		
Db 2557	TGTACACCGCTGTGGACAACAGATATGTGGATGTAGATTCGTGAGAGAACTTCTGTATATG	2616		
QY 1194	gtagtcctcgtgtgcctcagtggtccacgctgataagcgcaactgtccatatatgaagccagag	1253		
Db 2617	-----CGTGTGCCCATTTGCAAACTTGTCAAGAGATCTCAACACCGAAT	2660		
QY 1254	attgcttccttgytgaanaacttacagacagcttcataagacatgaagtacacaa	1313		
Db 2661	ATTGTATTTCTTTGGAGAGAGATCTCGGAAGAGAAATTTTCATCAACATGTTCACAAAGACAAG	2720		
QY 1314	gatgaagttgaccctccatgttatgtgatctctctgaaagtgagaccagtaagacta	1373		
Db 2721	CATAAAGTAGATCTTATTTAGTAGATGTGGATCTTCTCTAGAAGTTCCACCATGTTGCATTA	2780		
QY 1374	attccaagttctatacccacatgaagtgcctcaatatataatgaaggaaccttgcct	1433		
Db 2781	ATTCCTCATTTGGTGTGAATGAAGATTTCCACCAATTCATCAACAGCAAACTCCCTCCAA	2840		
QY 1434	catcaacatttgatgatgtagcctcttgagacgtcgatctgttaataatgaagtgtgt	1493		
Db 2841	CATTATTAATGCCGATATTTGAATTTGCTCGGAATTTGTGATGGCACCATCAGAGATATCTGT	2900		
QY 1494	cataggctagtgtygtgaatatgcacaact	1522		
Db 2901	TTCAGTCTTTGGAGGCTCTTCACGTGAAGT	2929		
RESULT 15				
LOCUS CAC49C4	35338 bp	DNA		
DEFINITION C.albicans cosmid Ca49C4.			PLN	10-NOV-1998
ACCESSION AL033503				
VERSION AL033503.1	GI:3859673			
KEYWORDS 5-amino-levulinic acid synthase; DNA mismatch repair protein; hem1; mitochondrial carrier protein; mlt5 family; nadh-ubiquinone oxidoreductase; opti pseudogene; peptide transport protein; phenylalanyl-tRNA synthetase; phars; ptc2; ras-related protein; sec4; transcription regulatory protein; transmembrane protein; vacuolar ATP synthase subunit E; vma4; wv domain.				
SOURCE ORGANISM Candida albicans.				
REFERENCE 1 (bases 1 to 35338)				
AUTHORS Tait,E., Simon,M.C., King,S., Brown,A.J., Gow,N.A. and Shaw,D.J.				
TITLE A Candida albicans genome project: cosmid contigs, physical mapping, and gene isolation				
JOURNAL Fungal Genet. Biol. 21 (3), 308-314 (1997)				
MEDLINE 97435544				
REMARK Article No. FG970983				
REFERENCE 2 (bases 1 to 35338)				
AUTHORS Murphy,L. and Harris,D.				

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
3 (bases 1 to 35338)
Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (09-NOV-1998) On behalf of the pilot sequencing project
on the Candida albicans strain 1161 genome. Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
barrell@sanger.ac.uk Cosmids supplied by Prof. Duncan Shaw, [3]
Department of Molecular and Cell Biology, The Institute of Medical
Science, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD,
UK

COMMENT

Notes:
Funding: sequencing funded by Becoulif Genomics Ltd. CDS are
numbered using the following system eg CAC20C1.01C. CA (C.
albicans), C20C1 (cosmid name),
.01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE/PIR
database are also included but some of these may be fortuitous.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.

FEATURES
source

1..35338
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="R"
/clone="cosmid Ca49C4"
283..993
/gene="Ca49C4.01"
283..993
/note="Ca49C4.01", len: 236 aa, unknown, repetitive.
transmembrane, similar eg to S. cerevisiae to YFB0, YEAST
hypothetical protein YFL010C (211 aa), fasta scores, opt:
334 z-score: 292.5 E(): 4.6e-09 (38.3% identity in 248 aa
overlap), contains PS01159 WW /rps5 /MWP domain signature,
Pfam match to entry PF00397 WW_rps5_MWP, WW domain"
/codon_start=1
/transl_table=12
/product="hypothetical transmembrane protein"
/protein_id="CA422011.1"
/db_xref="GI:3859674"
/translation="MSKDDSKSPQVDPGWAKFDEEYSTWYVDLTKKSGMDAPAG
TRFDSKSGDDVPVPPAYSPTEKSRNAPSNQPNRPAASNPQDQPGGQOYGO
GQPGGQOYGOOQPGGQOYGOOQPGGQOYGOOQPGGQOYGOOQPGGQOYGO
YGOOQPGGQOYGOOQPGGQOYGOOQPGGQOYGOOQPGGQOYGOOQPGGQOYGO
GYDNGGQYDGGDFGGDFP"

misc_feature
316..408
/gene="Ca49C4.01"
/note="Pfam match to entry PF00397 WW_rps5_MWP, WW domain,
score 44.30, E-value 1.2e-10"
328..408
/gene="Ca49C4.01"
/note="PS01159 WW /rps5 /MWP domain signature"
1276..1695
/gene="Ca49C4.02"
1276..1695
/gene="Ca49C4.02"
/note="Ca49C4.02", len: 139 aa, nadh-ubiquinone
oxidoreductase, similar to eg, NUPM_CAEEL putative
nadh-ubiquinone oxidoreductase 17.3 kDa subunit (EC
1.6.5.3) (150 aa) fasta scores, opt: 243, E(): 1.7e-12,
(37.1% identity in 105 aa overlap)"
/codon_start=1
/transl_table=12
/product="nadh-ubiquinone oxidoreductase"
/protein_id="CA422012.1"
/db_xref="GI:3859675"
/translation="MRETPILRQIPKIKLSVYLKADGVPPTGLAGIYKHPNPALV
TLNETLTKLENPADSVYRQSTFALTNKPKLVESEIENIENIENIGGLLEIYI
OAADELISALQGLKVMELERKLDQWVYFGKKI"

misc_feature
316..408
/gene="Ca49C4.01"
/note="Pfam match to entry PF00397 WW_rps5_MWP, WW domain,
score 44.30, E-value 1.2e-10"
328..408
/gene="Ca49C4.01"
/note="PS01159 WW /rps5 /MWP domain signature"
1276..1695
/gene="Ca49C4.02"
1276..1695
/gene="Ca49C4.02"
/note="Ca49C4.02", len: 139 aa, nadh-ubiquinone
oxidoreductase, similar to eg, NUPM_CAEEL putative
nadh-ubiquinone oxidoreductase 17.3 kDa subunit (EC
1.6.5.3) (150 aa) fasta scores, opt: 243, E(): 1.7e-12,
(37.1% identity in 105 aa overlap)"
/codon_start=1
/transl_table=12
/product="nadh-ubiquinone oxidoreductase"
/protein_id="CA422012.1"
/db_xref="GI:3859675"
/translation="MRETPILRQIPKIKLSVYLKADGVPPTGLAGIYKHPNPALV
TLNETLTKLENPADSVYRQSTFALTNKPKLVESEIENIENIENIGGLLEIYI
OAADELISALQGLKVMELERKLDQWVYFGKKI"

misc_feature

complement(2048..2638)
/note="Pfam match to entry PF00071 ras, Ras family, score
338.70, E-value 6.6e-98"
complement(2051..2683)
/gene="sec4"
complement(2051..2683)
/gene="sec4"
complement(2051..2683)
/note="Ca49C4.03c, len 210 aa, sec4, identical to O14462
SEC4P, contains PS00017 ATP /GTP-binding site motif A,
Pfam match to entry PF00071 ras, Ras family"
/codon_start=1
/transl_table=12
/product="ras-related protein sec4p"
/protein_id="CA422013.1"
/db_xref="GI:3859676"

gene

CDS

misc_feature

complement(3492..5270)
/gene="Ca49C4.04c"
complement(3492..5270)
/gene="Ca49C4.04c"
/note="Ca49C4.04c.1, phers, len: 592 aa, (branch site
upstream may indicate splicing) nearly identical to to
E1240164 O13432, phenylalanyl-L-rna synthetase (593 aa),
fasta scores, opt: 3785, (97.5% identity in 593 aa
overlap), also similar to S. cerevisiae SYR4_YEAST
phenylalanyl-L-rna synthetase (EC 6.1.1.20) FRSI OR YLR060W
(594 aa), fasta scores, opt: 2568, (62.8% identity in 596
aa overlap), contains PS00017 ATP /GTP-binding site motif
A"
/codon_start=1
/transl_table=12
/product="phenylalanyl-L-rna synthetase"
/protein_id="CA422014.1"
/db_xref="GI:3859677"

gene

CDS

misc_feature

misc_feature

complement(3783..3806)
/gene="Ca49C4.04c"
/note="PS00017 ATP /GTP-binding site motif A"
complement(5285..5291)
/note="tactaac, possible splice branch site for Ca49C4.04"
complement(5533..8253)
/gene="Ca49C4.05c"
complement(5533..8253)
/gene="Ca49C4.05c"
/note="Ca49C4.05c, len: 906 aa, similar to S. cerevisiae
D1009845 YFL034W (1073 aa), fasta scores, opt: 1895,
(41.4% identity in 929 aa overlap)"
/codon_start=1
/transl_table=12
/product="conserved hypothetical protein"
/protein_id="CA422015.1"
/db_xref="GI:3859678"

misc_feature

misc_feature

gene

CDS

complement(5533..8253)
/gene="Ca49C4.05c"
complement(5533..8253)
/gene="Ca49C4.05c"
/note="Ca49C4.05c, len: 906 aa, similar to S. cerevisiae
D1009845 YFL034W (1073 aa), fasta scores, opt: 1895,
(41.4% identity in 929 aa overlap)"
/codon_start=1
/transl_table=12
/product="conserved hypothetical protein"
/protein_id="CA422015.1"
/db_xref="GI:3859678"
/translation="MSSNRNSFDRPLRYWMSIGSRESIISPTHEGYSKPSKTSVTH
LPSLPYQATTSPESSLDNDEEVGAVDDATIEDLSSPNRNTSSFNRTTPRVH
SSSISTPIQEHSGSLVDKRPQDECELEKRDDEDFPSSGSMKMTISMDVYN
EKGLERNSKSGKDFIVDKNQFGYITIDEQVKNAAALDQKDFLFRPGEQORNN
AKYHKDIFLEQDDSDYDEDEDEISGELTLETMKMLDSQGFAYIGIAGLITVYMAVD

FT	variation	/label- Polymorphic_site
FT		9600
FT		/tag- r
FT	variation	/label- Polymorphic_site
FT		9951
FT		/tag- s
FT	variation	/label- Polymorphic_site
FT		1011
FT		/tag- t
FT	variation	/label- Polymorphic_site
FT		11250
FT		/tag- u
FT	variation	/label- Polymorphic_site
FT		12300
FT		/tag- v
FT	variation	/label- Polymorphic_site
FT		14007
FT		/tag- w
FT	variation	/label- Polymorphic_site
FT		15355..15360
FT		/tag- x
FT	polyA_signal	

PN MO9211387-A.
 PD 09-JUL-1992.
 PE 20-DEC-1991; 91WO-CA00457.
 PR 21-DEC-1990; 90GB-0027869.
 PR 20-MAY-1991; 91GB-0010865.
 PR 09-SEP-1991; 91GB-0019250.
 PA (UNGU-) UNIV GUELPH.
 PA (UTOR) UNIV TORONTO INNOVATIONS FOUND.
 P1 MacIennan DH, O'Brien PJ;
 DR MPI: 1992-250106/30.
 DR P-PSDB; R25450.
 XX
 PT Purified DNA mol. for diagnosis of porcine malignant hyperthermia
 PT -compiles DNA sequence encoding normal or mutant ryanodine
 PT receptor with specified endonuclease restriction map
 XX
 PS Disclosure: Fig 2; 96pp; English.
 XX
 CC The sequence given is the mutant pig ryanodine receptor (RYR1) gene
 CC from swine cDNA. The polymorphic sites were observed in comparisons
 CC of Pietrain and Yorkshire breeds. There are 17 polymorphisms between
 CC the two breeds. The polymorphism at position 1972 causes a mutation
 CC from Arg to Cys and this is thought to be the molecular basis of
 CC porcine malignant hyperthermia (MH). This mutation lies within the
 CC region of RYR1 that is concerned with the binding of regulators of Ca2
 CC release channel gating. Analysis of surrounding sequences suggests
 CC that this mutation lies within a beta strand domain comprising roughly
 CC of amino acids 520 to 830. RYR1 is the calcium release channel of the
 CC sarcoplasmic reticulum and is a large protein which spans the gap
 CC between the transverse tubule and the sarcoplasmic reticulum. The
 CC channel is activated by ATP, calcium, caffeine, and micro-molar
 CC ryanodine. It is inhibited by ruthenium red, tetracaine, calmodulin,
 CC high Mg2+ and ryanodine.
 CC
 CC Sequence 15377 BP; 3197 A; 4630 C; 4755 G; 2774 T; 21 other;

```

Query Match      1.9%   Score 76.2   DB 13:   Length 15377;
Best Local Similarity 49.2%   Pred No. 2.3e-05;
Matches 192; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY  51  ggcgaagagtgatgcgccttcgaagcgcgcgcgtcccttcgcgagcgccgacatg 110
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
db   12934  gagggcgcgagaaagagcgcgcgtctggggcccccggagcgcgcgagcgacccgcgcgcgcgc 12993

```

[illegible]

ID	Accession	Description
RESULT 9	V62176/C	V62176 standard; DNA; 117213 BP.
XX	V62176;	
XX	13-JAN-1999	(first entry)
DE	HSV-2 strain SB5	Contig ID 15 DNA sequence.
XX	HSV-2 strain SB5;	immunological response induction; therapy
KW	antiviral identification;	viral protein inhibitor; ss.
RK	Herpes simplex virus type 2.	
XX		
XX	Key	Location/Qualifiers
FH	CDS	755..1297
FH	/tag= a	
FT	/product= "ORF#1 protein"	
FT	/note= "encoded protein shown in W72170"	
FT	1170..2174	
FT	/tag= b	
FT	/product= "ORF#2 protein"	
FT	/note= "encoded protein shown in W72171"	
FT	2229..2930	
FT	/tag= c	
FT	/product= "ORF#3 protein"	
FT	/note= "encoded protein shown in W72172"	
FT	complement (3130..3735)	
FT	/tag= d	
FT	/product= "ORF#4 protein"	
FT	/note= "encoded protein shown in W72173"	
FT	complement (3802..6447)	
FT	/tag= e	
FT	/product= "ORF#5 protein"	
FT	/note= "encoded protein shown in W72174"	
FT	6017..8482	
FT	/tag= f	
FT	/product= "ORF#6g protein"	
FT	/note= "encoded protein shown in W72250"	
FT	6026..8482	
FT	/tag= g	
FT	/product= "ORF#6f protein"	
FT	/note= "encoded protein shown in W72249"	
FT	6065..8482	
FT	/tag= h	
FT	/product= "ORF#6e protein"	
FT	/note= "encoded protein shown in W72248"	

```
FT CDS 6167..8482
FT /tag= i
FT /product= "ORF#6d protein"
FT /note= "encoded protein shown in W72247"
FT CDS 6296..8482
FT /tag= j
FT /product= "ORF#6c protein"
FT /note= "encoded protein shown in W72246"
FT CDS 6326..8482
FT /tag= k
FT /product= "ORF#6b protein"
FT /note= "encoded protein shown in W72245"
FT CDS 6446..8482
FT /tag= l
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481..7486, aa: Ile)
FT /note= "encoded protein shown in W72244"
FT CDS 8457..9347
FT /tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in W72175"
FT CDS complement (9604..11855)
FT /tag= n
FT /transl_except= (pos: 11635..11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in W72176"
FT CDS complement (11905..14508)
FT /tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in W72222"
FT CDS complement (11905..14520)
FT /tag= p
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in W72223"
FT CDS 14399..15802
FT /tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in W72177"
FT CDS complement (15996..16286)
FT /tag= r
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in W72178"
FT CDS complement (16202..18064)
FT /tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in W72179"
FT CDS complement (18105..19661)
FT /tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in W72180"
FT CDS complement (19415..20074)
FT /tag= u
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in W72181"
FT CDS 20155..21453
FT /tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in W72182"
FT CDS complement (21326..22291)
FT /tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in W72183"
FT CDS complement (22546..24654)
FT /tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in W72184"
FT CDS 24684..25955
FT /tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in W72185"
FT CDS complement (26295..27251)
FT /tag= z

FT CDS /product= "ORF#19 protein"
FT /note= "encoded protein shown in W72186"
FT CDS complement (27630..31754)
FT /tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in W72225"
FT CDS complement (27630..31784)
FT /tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in W72224"
FT CDS complement (32067..33735)
FT /tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in W72187"
FT CDS 33140..34984
FT /tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in W72226"
FT CDS 33386..34984
FT /tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in W72227"
FT CDS complement (35205..37721)
FT /tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in W72188"
FT CDS complement (38058..39188)
FT /tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in W72189"
FT CDS 39090..39935
FT /tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in W72190"
FT CDS 40216..41973
FT /tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in W72191"
FT CDS 42206..44178
FT /tag= aj
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in W72192"
FT CDS complement (44853..47297)
FT /tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in W72193"
FT CDS 47122..47338
FT /tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in W72194"
FT CDS complement (47305..49662)
FT /tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in W72195"
FT CDS complement (50035..51666)
FT /tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in W72196"
FT CDS complement (51701..53575)
FT /tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in W72197"
FT CDS 54393..58115
FT /tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in W72198"
FT CDS complement (58060..58977)
FT /tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in W72199"
FT CDS complement (58970..60760)
FT /tag= ar
```

[illegible]

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Larson G, Drmanac R, Ctkvenjakov R, Dickson M, Drmanac S, Labat I;
XX Leschowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
DR WPI: 2000-126369/11.
XX
XX polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1, Page 995-996; 1097pp; English.
XX
CC A00010 to A02716 represent polynucleotides isolated from cDNA libraries
CC constructed from human colon cancer cell lines. The present invention
CC also describes a method of detecting differentially expressed genes
CC correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 1218 BP; 9 A; 31 C; 494 G; 37 T; 647 other;

[illegible]

XX X53491 standard; DNA: 114955 BP.
AC X53491;
XX
XX
DE 05-JUL-1999 (first entry)
XX
XX Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
XX Antisense oligonucleotide; multiple target; antisense treatment;
KW Impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
KW respiratory distress syndrome; pain; cystic fibrosis;
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
XX
XX Synthetic.
OS
XX W09913886-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 17-SEP-1998; 98WO-US19419.
PE
XX
XX 09-JUN-1998; 98US-0093972.
PR
XX 17-SEP-1997; 97US-0059160.
XX
XX (UVEC-) UNIV EAST CAROLINA.
PA
XX
XX Nyce JW;
PI
XX
XX WPI: 1999-223400/19.
DR
XX
XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT vasoconstriction
XX
XX Disclosure: Page 37; 120pp; English.
PS
XX
XX The specification describes antisense oligonucleotides (X52869-X55271)
CC directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene
CC initiation codons, genomic flanking regions, intron-exon borders, the
CC 5'-end, the 3'-end and the junction between coding and non-coding
CC regions and all segments of RNAs encoding proteins associated with one
CC or more diseases, conditions or mixtures. The antisense oligonucleotides
CC may be derived from sequences X55272-74. These multiple target
CC oligonucleotides (specifically X55180-271) can be used for the antisense
CC treatment of diseases and conditions. Typical diseases and conditions
CC are those associated with impaired respiration and inflammation,
CC including lung diseases, pulmonary vasoconstriction and inflammation,
CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC metastases, as well as all types of cancers which may metastasize or have
CC metastasized to the lungs, including breast and prostate cancer.
XX
XX
XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match	1.8%	Score 71.4	DB 20	Length 114955
Best Local Similarity	35.8%	Pred. No. 0.0001		
Matches 167	Conservative 33	Mismatches 266	Indels 1	Gaps 1

Qy	6	gcacagagagcgcagagcgcgagagcagagacatgtgaagatgcgcgacagagtgccg	65
Db 110581		gcggggcgccggcggnnnnnnssggccggcgccggcggnnnnnnssggcgccggcgccggcg	110522

[illegible]

C	1	79.4	2.1	1931	3	US-09-130-114-2	Sequence 2, Appl 1
C	12	76.8	2.0	152331	3	US-09-128-155-16	Sequence 16, Appl 1
C	3	73.2	1.9	15378	3	US-08-785-420-1	Sequence 1, Appl 1
C	4	70.2	1.8	1548	2	US-08-762-106-5	Sequence 5, Appl 1
C	5	70.2	1.8	1581	2	US-08-762-106-6	Sequence 6, Appl 1
C	6	69.8	1.8	8438	1	US-07-945-283-1	Sequence 1, Appl 1
C	7	69.6	1.8	12001	1	US-08-458-568A-11	Sequence 11, Appl 1
C	8	68.6	1.8	4524	2	US-08-845-998-7	Sequence 7, Appl 1
C	9	68.6	1.8	4524	3	US-09-206-537-7	Sequence 7, Appl 1
C	10	66.8	1.7	6530	3	US-08-146-930-1	Sequence 1, Appl 1
C	11	66.8	1.7	6530	3	US-08-458-240-1	Sequence 1, Appl 1
C	12	66.8	1.7	6530	4	PCT-US93-03993-1	Sequence 1, Appl 1
C	13	66.4	1.7	2580	3	US-09-030-863-2	Sequence 2, Appl 1
C	14	66.4	1.7	3452	2	US-09-130-114-1	Sequence 1, Appl 1
C	15	66.4	1.7	10596	1	US-07-884-811-15	Sequence 15, Appl 1
C	16	66.4	1.7	10596	1	US-07-885-971-15	Sequence 15, Appl 1
C	17	66.4	1.7	10596	1	US-08-087-783A-15	Sequence 15, Appl 1
C	18	66.4	1.7	10596	1	US-08-194-088B-15	Sequence 15, Appl 1
C	19	66.4	1.7	10596	2	US-08-194-087-15	Sequence 15, Appl 1
C	20	66.4	1.7	10596	4	PCT-US93-04648-15	Sequence 15, Appl 1
C	21	65.4	1.7	2823	1	US-08-398-008A-1	Sequence 1, Appl 1
C	22	65.4	1.7	2823	2	US-08-893-333-1	Sequence 1, Appl 1
C	23	64.6	1.7	9046	1	US-08-227-536-1	Sequence 1, Appl 1
C	24	64.6	1.7	9046	4	PCT-US93-04682-1	Sequence 1, Appl 1
C	25	64	1.7	2634	3	US-08-941-936-1	Sequence 1, Appl 1
C	26	63.8	1.6	954	3	US-08-680-506-5	Sequence 5, Appl 1
C	27	63.8	1.6	1347	3	US-08-680-506-8	Sequence 8, Appl 1
C	28	63.8	1.6	1587	3	US-08-600-506-6	Sequence 6, Appl 1

C 29	63.8	1.6	2233	3	US-08-680-506-4	Sequence 4, Appl.1
C 30	63.8	1.6	5037	2	US-08-680-506-2	Sequence 2, Appl.1
C 31	63.6	1.6	4257	2	US-08-690-473-1	Sequence 1, Appl.1
C 32	63.6	1.6	12001	1	US-08-458-5684-11	Sequence 11, Appl.1
C 33	61.8	1.6	1035	1	US-07-601-094-30	Sequence 30, Appl.1
C 34	61.8	1.6	1035	1	US-08-012-735-30	Sequence 30, Appl.1
C 35	61.8	1.6	1914	1	US-07-601-094-1	Sequence 1, Appl.1
C 36	61.8	1.6	1914	1	US-08-012-735-1	Sequence 1, Appl.1
C 37	61.6	1.6	4257	2	US-08-690-473-1	Sequence 1, Appl.1
C 38	61.4	1.6	530	3	US-08-758-662-4	Sequence 4, Appl.1
C 39	61.4	1.6	2833	1	US-08-398-0084-1	Sequence 1, Appl.1
C 40	61.4	1.6	2833	2	US-08-893-833-1	Sequence 1, Appl.1
C 41	61.4	1.6	2936	2	US-08-714-667-10	Sequence 10, Appl.1
C 42	61.4	1.6	2936	2	US-08-393-540-10	Sequence 10, Appl.1
C 43	61.4	1.6	2936	2	US-08-714-537-10	Sequence 10, Appl.1
C 44	61.2	1.6	2731	5	5215881-2	Patent No. 5215881
C 45	61.2	1.6	44377	2	US-08-604-2270-7	Sequence 7, Appl.1

ALIGNMENTS

```

RESULT 1
US-09-130-114-2/c
: Sequence 2, Application US/09130114
: Patent No. 5976807
:
GENERAL INFORMATION:
:
APPLICANT: HOTLICK, Robert A.
APPLICANT: DamaJ, Bassam B.
APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: TITLE OF INVENTION: From Multiple Transfected Episomes
: FILE REFERENCE: 0867/1D9030U1
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ. ID NOS.: 36
:
SOFTWARE: fastSeq for Windows Version 3.0
:
SEQ ID NO 2
:
LENGTH: 1931
:
TYPE: DNA
:
ORGANISM: EBNA
:
US-09-130-114-2

```

Query Match	2.1%;	Score 79.4;	DB 2;	Length 1931;
Best Local Similarity	49.5%;	Pred. No. 2.1e-08;		
Matches 237;	Conservative	0;	Mismatches 236;	Indels 6; Gaps 1

[illegible]

[illegible]

RESULT 4
US-08-762-106-5
; Sequence 5, Application US/08762106

```

1  TITLE OF INVENTION:  READING FRAME INDEPENDENT EPITOPES
2  TITLE OF INVENTION:  TAGGING
3  NUMBER OF SEQUENCES:  47
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Harits Brotman
6  STREET:  202 Coast Blvd., Suite 111
7  CITY:  La Jolla
8  STATE:  California
9  COUNTRY:  US
10 ZIP:  92037
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE:  Floppy disk
13 COMPUTER:  IBM PC compatible
14 OPERATING SYSTEM:  PC-DOS/MS-DOS
15 SOFTWARE:  PatentIn Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER:  US/08/762,106
18 FILING DATE:  09-DEC-1996
19 CLASSIFICATION:  435
20 ATTORNEY/AGENT INFORMATION:
21 NAME:  Brotman, Harits F.
22 REGISTRATION NUMBER:  35,461
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE:  (619) 654-2428
25 INFORMATION FOR SEQ ID NO:  5:
26 SEQUENCE CHARACTERISTICS:
27 .. LENGTH:  1548 base pairs
28 .. TYPE:  nucleic acid
29 .. STRANDEDNESS:  single
30 .. TOPOLOGY:  linear
31 MOLECULE TYPE:  cDNA
32 HYPOTHETICAL:  NO
33 ANTI-SENSE:  NO
34 US-08-762-106-5

```

Query Match	1.88;	Score 70.2;	DB 2;	Length 1548;
Best Local Similarity	47.88;	Pred. No. 2e-06;		
Matches 204; Conservative	0;	Mismatches 223;	Indels 0;	Gaps 0

[illegible]

Oy	189	g g c c g a g c c g a a c g a g c a a t g y g g c c g g c g g c g g g t g t g a g c g a g c g c g c g	248
Db	1093	g c g t g c g a c g a c c c c a t t g c c c t t c a a g c c g c g c c c c g a c t a c g c c a g c g a g	1152
Oy	249	g c c c c g y c g c t g t g y c g a g a g c g g c a g y g c g c g a g c y c g a g c y c g a g y g a g c c	308
Db	1153	g a g c g c g c c a a c g c c a a g t a g g a g c g c g a a a m a g c g a g c g c g a g a m a g c c g c g c c	1212
Oy	309	c c g g c g a c g a g c c t f g a c c g g a g a c g y a g a a e t g g t c c g c c t c g g g a g a g c c a g g	368
Db	1213	g a g c g c g a g c g c g c g c g c g a g a c a a m a g c c t c t g a g a g a g t t g c c g c c a c g c g c	1272
Oy	369	g c g g c t g a c g a c t t c g a c g a c g a c a g g c y a g a g a g a g a c a g a c y c g c y c g a c g	428
Db	1273	g c g c g c g a g c g c g a g a g c c g c c g c a a g a g c c c c g c g a g c c c g t c g t c c g c g a t	1332
Oy	429	g c g g c g g	435
Db	1333	g c c t g c g	1339

RESULT 5
US-08-762-106-6
: Sequence 6, Application US/08762106

APPLICANT: Jarvik, Jonathan W.
 TITLE OF INVENTION: READING FRAME INDEPENDENT EPIPTOPES
 TITLE OF INVENTION: TAGGING
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Harris Brotman
 STREET: 202 Coast Blvd., Suite 111
 CITY: La Jolla
 STATE: California
 COUNTRY: US
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/762,106
 FILING DATE: 09-DEC-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Brotman, Harris F.
 REGISTRATION NUMBER: 35,461
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 654-2428.
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1581 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-762-106-6

Query Match	1.8%;	Score 70.2;	DB 2;	Length 1581;
Best Local Similarity	47.8%;	Pred. No. 2e-06;		
Matches 204;	Conservative	0;	Mismatches 223;	Indels 0;
			Gaps	0

[illegible]

Db	1006	ACGGGCGACAGGTGGAGCAGCCTTGCGGCGAGCAGCGCAGCCAAAGTAGAGCGAGAGCTGCA	1065
QY	129	gcgagacagagccgctccgcgaagagagccccgcgcagagacgagcctggtcgagcccg	188
Db	1066	GCGAAGGACGTGGAGGCGCGTGGCGGCGCGCCACACGTTTGCTGTGGCGACGTTCAAGCC	1129
QY	189	ggcgagaccgagacgagacagctgtagcgccgagccgagcggtgtgtgagcgcgagcgccg	248
Db	1126	CGGTTGCGAGCCGCCGATGCCGTTGAGAGCGCGCGCGCGAGAGCTGACGGCGGAGCGGAG	1185
QY	249	gccccgcgagcgctgtgagagagcgagcgagcgcgagcgagcgagcgagcgagcgcc	308
Db	1186	GAGCGCGCCAAACGCCAGGTGGAGGCGCGAACAAGCGGAGCGCGGAGAGAGCGCGCC	1245
QY	309	ccgagcgagcgagcgctgtgagcgagagcgagacaaatgtggtccgacctgagcgagagcgagag	368
Db	1246	GAGCGCGGAGCGCGCGCGCGGAGAGCAAGAAAGCGCTGCTGAGAGAGTTGGCGGCGCACGCG	1305
QY	369	ggcgagctgacgaacttcgagcagacagacagagagcgagagagagagacgagagcgagcgagcg	428
Db	1306	GGCGGGGAGGCGGAGAGCGCGCGCGAGAGACCGCCGCGGAGACCGCGCTGCTGCCGAGT	1365
QY	429	gagcgagcg 435	
Db	1366	GCGCTGG 1372	

RESULT 6
US-07-945-283-1
: Sequence 1, Application US/07945283

```

1 GENERAL INFORMATION:
2 APPLICANT: Cheung, Andrew K.
3 APPLICANT: Wesley, Ronald D.
4 TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
5 TITLE OF INVENTION: Involving The Ep0 and Lf1 Genes
6 NUMBER OF SEQUENCES: 7
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Curtis P. Ribando
9 STREET: 1815 No. 5352596th University Street
10 CITY: Peoria
11 STATE: IL
12 COUNTRY: USA
13 ZIP: 61604
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patent Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/07/945, 283
21 FILING DATE: 19920911
22 CLASSIFICATION: 424
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Ribando, Curtis P
25 REGISTRATION NUMBER: 27976
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 309-685-4011 ext.513
28 TELEFAX: 309-685-4128
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 8438 base pairs
32 TYPE: NUCLEIC ACID
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 HYPOTHETICAL: NO
37 ANTI-SENSE: NO
38 ORIGINAL SOURCE:
39 ORGANISM: Pseudorabies virus
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 622..6495

```

```

?      FEATURE:
?      NAME/KEY:      variation
?      LOCATION:      replace(1099, "g")
?      FEATURE:
?      NAME/KEY:      variation
?      LOCATION:      replace(1267, "t")
?      FEATURE:
?      NAME/KEY:      variation
?      LOCATION:      replace(1381, "c")
?      FEATURE:
?      NAME/KEY:      variation
?      LOCATION:      replace(1566, "c")
?      FEATURE:
?      NAME/KEY:      variation
?      LOCATION:      replace(7010, "g")
US-07-945-283-1

```

Query Match	1.88;	Score 69.8;	DB 1;	Length 8438;
Best Local Similarity	.54.58;	Pred. No. 5.1e-06;		
Matches 140;	Conservative 0;	Mismatches 117;	Indels 0;	Gaps 0;

[illegible]

RESULT 7
 US-08-458-11/C
 : Sequence 11, Application US/08458568A
 : Patent No. 5821339
 : GENERAL INFORMATION:
 : APPLICANT: Schaffer, Priscilla A.
 : APPLICANT: yeh, Lily
 : TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
 : TITLE OF INVENTION: Infections
 : NUMBER OF SEQUENCES: 15
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Woodcock, Mashburn, Kurtz, Mackiewicz & No. 5821339.r1s
 : STREET: One Liberty Place, 46th floor
 : CITY: Philadelphia
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Wordperfect 5.1
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/458,568A
 : FILING DATE: 02-JUNE-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/065,146
 : FILING DATE: 05-MAY-1993
 : CLASSIFICATION: 435

RESULT 9
 US-09-206-537-7
 : Sequence 7, Application US/09206537
 : Patent No. 6130052
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Van Baren, Nicolas
 : APPLICANT: Coulie, Pierre G.
 : APPLICANT: De Smet, Charles
 : APPLICANT: Lucas, Sophie
 : APPLICANT: Boom, Thelery
 :
 : TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
 :
 : NUMBER OF SEQUENCES: 16

[illegible]

```

US-08-458-240-1
: RESULT 11
: Sequence 1, Application US/08458240
: Patent No. 6143727
: GENERAL INFORMATION:
:   APPLICANT: Roop, Dennis R.
:   APPLICANT: Rothnagel, Joseph A.
:   APPLICANT: Greenhalgh, David A.
:   TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
:   TITLE OF INVENTION: AND METHODS OF USE
:   NUMBER OF SEQUENCES: 4
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: LYON & LYON
:     STREET: 611 West Sixth Street
:     CITY: Los Angeles
:     STATE: California
:     COUNTRY: U.S.A.
:   ZIP: 90017
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
:   SOFTWARE: WordPerfect (Version 5.1)
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/458,240
:     FILING DATE:
:   CLASSIFICATION:
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US/08/146,930
:       FILING DATE:
:     APPLICATION NUMBER: 07/876,286
:       FILING DATE: April 30, 1992
:     APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
:       FILING DATE: October 29, 1993
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Watbuid, Richard J.
:     REGISTRATION NUMBER: 32,327
:     REFERENCE/DOCKET NUMBER: 204/152
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (213) 489-1600
:       TELEFAX: (213) 955-0440
:     TELER: 67-3510
:   INFORMATION FOR SEQ ID NO: 1:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 6530 bases
:       TYPE: nucleic acid
:       STRANDEDNESS: single
:       TOPOLOGY: linear
:     MOLECULE TYPE: DNA (genomic)
:   US-08-458-240-1

```

Query Match	1.7%	Score 66.8	DB 3	Length 6530
Best Local Similarity	48.9%	Pred. No. 2.1e-05		
Matches 179	Conservative 0	Mismatches 187	Indels 0	Gaps 0

Qy	9	gagagagcgagagcgagagagccagagagagcagctctgaagatattgcgcgacgagtgctgc	68
Db	3106	gcggcgccgcctctctgtctgtatgagcgcttcaatctactccgagagctgtatgacggcgccgctctca	3165

[illegible]

```

RESULT 12
PCT-US93-03993-1
; Sequence 1, Application PC/TUS9303993
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
; TITLE OF INVENTION: VECTOR SYSTEMS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03993
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6530 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US93-03993-1

```

Query Match	1.7%;	Score 66.8;	DB 4;	Length 6530
Best Local Similarity	48.9%;	Pred. No. 2.1e-05;		

QY 129 gccgacgagccgctccgcaagagagcccgccgagacgagcctgagcctcgagccg 188
Db 1679 GAGGGGCGAGAGGGGCGAGAGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 189 gccgagccgagcagcagctgagcagcagcagcagcagcagcagcagcagcag 248
Db 1619 GGGCAGGAGCAG 1560
QY 249 gcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 308
Db 1559 GAGCAG 1500
QY 309 ccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 368
Db 1499 CAGGAG 1440
QY 369 gccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 428
Db 1439 GCCGGGCTCGAG 1380
QY 429 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 460
Db 1379 GAGGTAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1348

RESULT 15

US-07-884-811-15
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/884,811
FILING DATE: 19920518
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-884-811-15

Query Match 1.7%; Score 66.4; DB 1; Length 10596;
Best Local Similarity 46.7%; Pred. NO. 3.1e-05;
Matches 211; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 9 gagagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 68
Db 2586 GAGGGGCGAGAGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2645
QY 69 gcccttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 128
Db 2646 GAGGGGCGAG 2705
QY 129 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 188
Db 2706 GAGGGGCGAG 2765
QY 189 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 248
Db 2766 GGGCAG 2825
QY 249 gcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 308
Db 2826 GAGCAG 2885
QY 309 ccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 368
Db 2886 CAGGAG 2945
QY 369 gccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 428
Db 2946 GCCGGGCTCGAG 3005
QY 429 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 460
Db 3006 GAGGTAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3037

Search completed: February 27, 2001, 06:01:27
Job time: 41919 sec

```
Db 1741 gtgtgtgtggaagaaacccaagaagtaacagctagtaagtaagtgtgagaacattaatg 1800
QY 1801 tggaaataccagaatttaagagctgtgtgttccgcgtactctcagacaaataatgaagaact 1860
Db 1801 tggaaataccagaatttaagagctgtgtgttccgcgtactctcagacaaataatgaagaact 1860
QY 1861 cagtttcagaaacagtgaaagaaatgtctgccttaataagacttgcaagagacatagta 1920
Db 1861 cagtttcagaaacagtgaaagaaatgtctgccttaataagacttgcaagagacatagta 1920
QY 1921 agcgggttgaaggtaatacctaactgtttgttaaccacaaatcgttacaatacttcacagcgt 1980
Db 1921 agcgggttgaaggtaatacctaactgtttgttaaccacaaatcgttacaatacttcacagcgt 1980
QY 1981 ctgaagtaatacctaactgtgaagatgaagctgttctccttaagttccctgtgagcaata 2040
Db 1981 ctgaagtaatacctaactgtgaagatgaagctgttctccttaagttccctgtgagcaata 2040
QY 2041 gtgacagtgagcacaatgcagagctccaaagtttagaagaaccccttggaagatgaaagtga 2100
Db 2041 gtgacagtgagcacaatgcagagctccaaagtttagaagaaccccttggaagatgaaagtga 2100
QY 2101 tgaagaatcttaacaaatgctgtgaagatgatacagagagccggaatgtgctgagagat 2160
Db 2101 tgaagaatcttaacaaatgctgtgaagatgatacagagagccggaatgtgctgagagat 2160
QY 2161 ctggaatttgagctgaatggaagagatcaagaagttgttaatgaagctatagctacaagac 2220
Db 2161 ctggaatttgagctgaatggaagagatcaagaagttgttaatgaagctatagctacaagac 2220
QY 2221 aggaatttgacagatgaaatacctacacacagaacataacacatttgaaagctgcagg 2280
Db 2221 aggaatttgacagatgaaatacctacacacagaacataacacatttgaaagctgcagg 2280
QY 2281 atccagaatttgctccacacagcaatggaaactttagacatgcaaaaaaagaaatgaaatg 2340
Db 2281 atccagaatttgctccacacagcaatggaaactttagacatgcaaaaaaagaaatgaaatg 2340
QY 2341 ctgtgaacttgaaacagaacatctgaagaatgtatcatatataagacttgaaatagaatg 2400
Db 2341 ctgtgaacttgaaacagaacatctgaagaatgtatcatatataagacttgaaatagaatg 2400
QY 2401 tctctcttgataattcttaaatgtccatacattctcgtttgttactgttaacttcaaacatg 2460
Db 2401 tctctcttgataattcttaaatgtccatacattctcgtttgttactgttaacttcaaacatg 2460
QY 2461 ttggttgaacttcaatctccttcaaggttcaattgtatgatacatctcgtatgtatata 2520
Db 2461 ttggttgaacttcaatctccttcaaggttcaattgtatgatacatctcgtatgtatata 2520
QY 2521 atttgttttttgcctaataagtttcaaaccttttaaaagtttcaaaagccattgaaatg 2580
Db 2521 atttgttttttgcctaataagtttcaaaccttttaaaagtttcaaaagccattgaaatg 2580
QY 2581 taatgaaagggaacagcctatctagaacaaagaaatgtaattccaacttttctgttctg 2640
Db 2581 taatgaaagggaacagcctatctagaacaaagaaatgtaattccaacttttctgttctg 2640
QY 2641 aacatgaaatagtttaaaagccccaattctcgttctcgtgaactttatatttaagaacag 2700
Db 2641 aacatgaaatagtttaaaagccccaattctcgttctcgtgaactttatatttaagaacag 2700
QY 2701 ttaactttttaaacaacttgacatlttccaanaactgtgagcagctaaacttttaaatca 2760
Db 2701 ttaactttttaaacaacttgacatlttccaanaactgtgagcagctaaacttttaaatca 2760
QY 2761 gatgacttgaatgtagagagtagcagacogtctcgtgagagacccaacttggcctcagt 2820
Db 2761 gatgacttgaatgtagagagtagcagacogtctcgtgagagacccaacttggcctcagt 2820
QY 2821 gtgtgaagcgtacttaactgcatcgttttctgttaactgtcagacgttgtaatgtccaac 2880
Db 2821 gtgtgaagcgtacttaactgcatcgttttctgttaactgtcagacgttgtaatgtccaac 2880
```

```
Db 2821 gtgtgaagcgtacttaactgcatcgttttctgttaactgtcagacgttgtaatgtccaac 2880
QY 2881 aggcctctgagacttaactgtataatgtaattggaatgtgttccagtgttccagaaca 2940
Db 2881 aggcctctgagacttaactgtataatgtaattggaatgtgttccagtgttccagaaca 2940
QY 2941 atagtctctgtctataagttcccttaagtttgaataattggcattgtttaaataac 3000
Db 2941 atagtctctgtctataagttcccttaagtttgaataattggcattgtttaaataac 3000
QY 3001 ctactactgtgtgagacgtgcataagatcttcaaccacaaatctgcgaagatgtgaat 3060
Db 3001 ctactactgtgtgagacgtgcataagatcttcaaccacaaatctgcgaagatgtgaat 3060
QY 3061 gcaaaagccttctgaaatcttaataatgttaacttactcgtggaagatgttaataatgtgac 3120
Db 3061 gcaaaagccttctgaaatcttaataatgttaacttactcgtggaagatgttaataatgtgac 3120
QY 3121 tgcgtgttttccataatgaggaagaaataggccttcttaataagttcccaagatcata 3180
Db 3121 tgcgtgttttccataatgaggaagaaataggccttcttaataagttcccaagatcata 3180
QY 3181 agataaatttgatctcaaccagaagtacactgttgcctgttgaagatlttggtataatg 3240
Db 3181 agataaatttgatctcaaccagaagtacactgttgcctgttgaagatlttggtataatg 3240
QY 3241 atcccaagtggttagcctgtatataatgagatgaatcagatccaatagccaatgaaac 3300
Db 3241 atcccaagtggttagcctgtatataatgagatgaatcagatccaatagccaatgaaac 3300
QY 3301 taagtcttaattttaaagagcttaagcttgccttcaaacaggagatccaatttctcaact 3360
Db 3301 taagtcttaattttaaagagcttaagcttgccttcaaacaggagatccaatttctcaact 3360
QY 3361 gccagaacttlttagccttcaaacagttcacacactcagaagacagatattatcttaacag 3420
Db 3361 gccagaacttlttagccttcaaacagttcacacactcagaagacagatattatcttaacag 3420
QY 3421 acttcttggaaacatgtgcccacaatttaaatctatcatgtggttttagatattatcaaa 3480
Db 3421 acttcttggaaacatgtgcccacaatttaaatctatcatgtggttttagatattatcaaa 3480
QY 3481 aaaaatgatttgaataatagctgttctttagaataaataacccaagttaagaccattact 3540
Db 3481 aaaaatgatttgaataatagctgttctttagaataaataacccaagttaagaccattact 3540
QY 3541 gccagaaggaagaaatgaatagtagctcatttccctacccaagaagataactgaaattatc 3600
Db 3541 gccagaaggaagaaatgaatagtagctcatttccctacccaagaagataactgaaattatc 3600
QY 3601 tggctacaactaaagaatgcagiatatattagtttccaattgtgcaatgaatgtgttgcata 3660
Db 3601 tggctacaactaaagaatgcagiatatattagtttccaattgtgcaatgaatgtgttgcata 3660
QY 3661 agacaatatatttgaataatgtttgaataatttttcaatgttcaatggaagactgttttc 3720
Db 3661 agacaatatatttgaataatgtttgaataatttttcaatgttcaatggaagactgttttc 3720
QY 3721 agctcttttataatgtgacataagacttttaagtaacttgacatgttggtagacgct 3780
Db 3721 agctcttttataatgtgacataagacttttcaatgttgaatgttggtagacgct 3780
QY 3781 taatgactgataatcttccccaacttttgaataatcaaaaacagttgttataactaaa 3840
Db 3781 taatgactgataatcttccccaacttttgaataatcaaaaacagttgttataactaaa 3840
QY 3841 aaaaaaaagtgcagcgcgcgcgaattc 3869
Db 3841 aaaaaaaagtgcagcgcgcgcgaattc 3869
```


Db 3892 ttccaccatgcagatggtttgtgtatcatgagatatttaattgaaagtgtttt 3891

Qy 3692 aaattattttacagtgagacgctttccagctcttttatactgtacatagaatttat 3751

Db 3892 aaattatttttaccagtgagacgctttccagctcttttatactgtacatagaatttat 3951

Qy 3752 gtaat---cggcgatagttttgttaacgctttaagacttgatattcttcccaactt 3808

Db 3952 gtaattaccgagcatagttttgttagactgtttaagactgtga-tacttccccaactt 4010

Qy 3809 ttgaaatcaaaaaacagtgctttataact 3836

Db 4011 ttgaaatcaaaaaacagtgcttttataact 4038

RESULT 3
US-09-572-411-6424/c
Sequence 6424, Application US/09572411
GENERAL INFORMATION:
APPLICANT: Gearling, David P.
APPLICANT: Holtzman, Douglas A.
APPLICANT: Xu, Yong Yao
APPLICANT: Acton, Susan L.
APPLICANT: Shyjan, Andrew W.
APPLICANT: Fraser, Christopher C.
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

```

CURRENT APPLICATION NUMBER: 05/09/5712,411
CURRENT FILING DATE: 2000-05-12
EARLIER APPLICATION NUMBER: 60/133,993
EARLIER FILING DATE: 1999-05-13
EARLIER APPLICATION NUMBER: 60/135,617
EARLIER FILING DATE: 1999-05-24
EARLIER APPLICATION NUMBER: 60/135,616
EARLIER FILING DATE: 1999-05-24
EARLIER APPLICATION NUMBER: 60/135,817
EARLIER FILING DATE: 1999-05-24
EARLIER APPLICATION NUMBER: 60/135,950
EARLIER FILING DATE: 1999-05-24
EARLIER APPLICATION NUMBER: 60/135,679
EARLIER FILING DATE: 1999-05-24
EARLIER APPLICATION NUMBER: 60/135,630
EARLIER FILING DATE: 1999-05-24
NUMBER OF SEQ ID NOS: 7417
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6424
LENGTH: 4326
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(4326)
OTHER INFORMATION: n = A,T,C or G

```

Query Match	53.9%	Score 2084.2;	DB 21;	Length 4326;
Best Local Similarity	77.7%	Pred. No. 0;		
Matches 3104; Conservative	0;	Mismatches 658;	Indels 235;	Gaps 39

0y	56	cgagtgtagcctcgcccttaagccgacgagccctcccttcgagc---	gscgcgcatgta	112
4302	CGAAGCGGGCCCTCGCCCTTAGCGCCGCGGCGCTCCCTTAGCGGGGGGCCGAGAGGGA			4243
0y	113	ggccgcgtcgacgcgacgagcgcgtccgaagagcccgccgagacggccgtg		172
4242	GCGCGGCTGCTCCCGCGCGGGAGCGCGCTCCGCAAGAGCGCGGAGAGTGTGTCGGG			4183
0y	173	ccttcggcgacgcgcggcgagcgcgagcgcgacagcgtggcg-----	cctggcggc	220
4182	CTCTGACGAGAGCCCGCTTTATTCTCGGCTGATAGGAGCCGACCAACGCTCTTGGCGCGG			4123

[illegible]

```
|||||
Db 3043 CATAGAGCCATGAGTATGACAAAGATGAAAGTTGACCTCCATGTTATTTGGGCTTCC 2984
Oy 1350 ctgaagtgagagccagtgacctaattccaaagttctatacccaatgaagtgccctcaata 1409
Db 2983 CTCAAAAGTAAGACCGTAGCTAAATTCGAGTTCATCCCATCCCATGAAAGTGGCTGAGATA 2924
Oy 1410 ttaataaaggaagcccttgctcctacatacttgatgtagagctcttggaagctgc 1469
Db 2923 TTAATTAATAGGAACCTTTGCTCTACATGCTATTTGATGAGAGCTTCTTGAGAGCTGT 2864
Oy 1470 gatgtataatgaagtgatgtgtcatagagctaggtgtgtaatatgccaactgtgtgt 1529
Db 2863 GATGTCATATTAATGAATGTGTCTAGTATGAGTGTGGAATGATGCCAACTTTGCTGT 2804
Oy 1530 aacctgtaaagcttccagaatactacgtaaaaacotccagcccaaaaaggaattgtgt 1589
Db 2803 AACCTGTAAACCTTTCAGAAATTAAGTGAACCTCCAGAAACAAAGAAATTTGGCT 2744
Oy 1590 cattatcagagttgcacacacactctctcatatttggaagactcaagttcaactga 1649
Db 2743 TATTTGTCAAGATTTGCCACCCACACCTCTTCAATGTTTCAGAAAGCTCAAGTTCCACGAA 2684
Oy 1650 agaactgtacccaagaactctctgtgtatgtctacactgttagaaccagaacaaacaac 1709
Db 2683 AGAAGTTCCACCAAGATTTCTTCAAGATGTCACACTTTTAGACCAAGCTAGAGT 2624
Oy 1710 aatgtataatgttagaagatcgtcaatcaa---gtgtgtggaagaaaaacccaaga 1766
Db 2623 AATG---ATGATTTTGAATGTGTCTGAATCAAAAGTTGTATGGAAGAAACACACGAA 2567
Oy 1767 gtagagactagtagaagatgttagaacattatg-----tggaaaaatccaatlttaag 1820
Db 2566 GTACAAACTCTTCAAGAAATGTTGAAATGATGCTGAACACATGGAATAATCCGATTTGAG 2507
Oy 1821 gctgtgtgtccagtagctcagacaaaatgtgaagaactcagctgtgcagaacagtga 1880
Db 2506 AATGTTGGTGTAGTACTGCGGAGAAATGAAGAAGACTTCACTGGCTGGACAGGAA 2447
Oy 1881 aaatgcgtgacctaaatagactgtcaaaagagcagatagtaagcggctgtgaaggtatcaa 1940
Db 2446 AAATCTGCGCTTAATAGAGTGGCAAAAGAGCAGATAGAGCGGCTGATGTATTCAG 2387
Oy 1941 tacctgtgtgtaccccaaatcgttaacataatccacggtgcgtgaggtataactcagact 2000
Db 2386 TATGCTGTTTTCACACCAATGTTACATTTTCCATGGCGGTGAGGTATATTCAGCTGT 2327
Oy 2001 gaagatgcagctgtctcctctagctgtgtgcagtaacagctgtgcacatgtccag 2060
Db 2326 GAAAGATGAGCTTTATCTCTAGTCTTGTGGCAGTAAAGATGATGTGGACATGCCAG 2267
Oy 2061 agtccaaagttgaagaaccccttggaagaatgaagtgaatgtgaagatcttacaatgtgc 2120
Db 2266 AGTCCAAAGTTTGAAGAAGCCATGTGAGATGAAGAATGAAGAATTTACAAATGGCC 2207
Oy 2121 ttggaagaatgataacggagagcgccgaatgtgtgtggaagatctcgtgtgtgagctgtga 2180
Db 2206 TTAGAGATGAGACCTGATGTTTCCAGAGAGAGCTGGAGAGCTGGATTTGGAGCTGATGGA 2147
Oy 2181 ggggagatcaagaggtgtgttaagctatagcttacaagaacaggaatgtgacatgtgaac 2240
Db 2146 GATGATCAAGAGGCAATTAATGAGCTATATCTGTGAACACAGAGACTAACACAGATGAGAC 2087
Oy 2241 tatcatcagacaaatcatalaactatltgaagctgtccgagatccagaaatgtctccaca 2300
Db 2086 TATTCATCAAAACAAATCAT---AGTGTAAATTAATTTGTCAGAGTAAAGAAATTTGTCCACA 2030
Oy 2301 gcaatggagactttgacatgtcaaaaaaatgaatgttctcgtgtgaac-----tgaaca 2356
Db 2029 GCATTAGGAACCTTAGCAGTGC---AAATGTAATGTTTACTTGTGAACCTGATAGAGCAA 1973
Oy 2357 ggaatctgaagaatgtattacttataagactgtg-aaatagatgtg-tctctgtgaataat 2414
|||||

Db 1972 GGAACACGAAAGGTGTAATATTATAGGTTGGTAATAATAGATGTTTTTCATGATAAAT 1913
Oy 2415 tctaaagttcaatcatcttctgtgttgaactgttaacttaaacatgtgtgtgact---- 2470
Db 1912 TTTTAATCTTATATTTTGTACTGTGTACAAACTCAACATCACTTTTGTGTTTTTAAAA 1853
Oy 2471 -----tcacttccttcaag 2487
Db 1852 AAAAAAGTACTAAGTATCTCAATCAGCTGTGTGTCAGAGCTAACTTTCTTTTAAAG 1793
Oy 2488 ttcattgtatgatacatcgtatgtatgtata---attgttcttgcctaaatgagc 2543
Db 1792 TTTATTTGTATGATAATATTCATATGTAATATAATTTTTTTGTTTGTGTAGAGACT 1733
Oy 2544 tcaacactttaaagtttc---aaaagcaatgtgaatg-----ttaaagtaagggaaca 2596
Db 1732 TTCACATTTTAAAGTTTTCAAAAGCCATGGAGATGTAAATTAATTAAGGAGACA 1673
Oy 2597 gctatcctagacaaaagaatgtgtatctcaacttcttctgttgaacatgtgaatgtta 2656
Db 1672 GCTAATCTAGACCAAGAAATGATATT---TCACTTTCTTTGTATCAATGATGATGTTG 1616
Oy 2657 aagccctcaattctgtctgtgaacttlatltttagagacagttlaacttlltaaacac 2716
Db 1615 AAGTACTCAAAATCTGTACGCTAAACCTTTGTATCTTTAACAACATTTTAAACAC 1556
Oy 2717 tggcaatccaaaactgtgtgagctaaactttaaactcacagatgactgttaatgtg 2776
Db 1555 TGGCATTTTCCAAAAC-TGTGGCAGCTAACTTTTAAATCTCAAAATGACATGCAAGTGTG 1497
Oy 2777 a-----ggagtcagcacagtgctgagacac----- 2804
Db 1496 AGTAGAAGAGTCAACAAATATGTGGGAGACACGCTGGTGTCTTACTTTTAAAGTA 1437
Oy 2805 -----aaaactgtggtcaggtgtgtgaagcgtactactgtcatcgtgtgt 2850
Db 1436 ATACTGTGTGAAGATTTCAAGATTAATTTATTTTACGTCAAAATGAAGATGGCTTTTG 1377
Oy 2851 taactgtgcagac--gtgttaatgtccaaaagagccctgtgactaatcgtataatga 2908
Db 1376 TACTTCTGTGACATGATGATGCTATTTGCTCATTAACATCACTGCAAAACAA 1317
Oy 2909 ct-----tggaaatgtgttcaagttgtgtctgaagaacaaatagtgctgtctatagtg 2962
Db 1316 ATTAATGCTTTGGAAATGTTTCAAGTGTGTGAAACATTAAGTGGCTGTGGAT---CC 1260
Oy 2963 cccttagttgaatattgtccattgttlaataactaactaacgtgtgtgaagcgtgc 3022
Db 1259 CTTTACTTTTGAATATTTTGCCATTTGTGTTTAAATCCTATCAGCTGTGTAGAGCTTGC 1200
Oy 3023 atagatcttcacacaaat-----actgccaagaatgtgaatalgtcaaaccttctgaat 3077
Db 1199 ATTGATCTTTTCCACAAAGTATTAACCTGCCAAATGTGAATATGCAAACTTTGCAAT 1140
Oy 3078 c--taataatgtactcttcaactgtggtggaagatgttaaatlttgagacgtcgttcttcact 3135
Db 1139 CTATTAATTAATGTGACTTCTTCACTGGGAGAGTGAATAATTTTGGACGCTG-TTTTCCAT 1081
Oy 3136 aatgaagaagacatagggcctctaat-taaagttcccaagtcataaga--taaatgtga 3192
Db 1080 AATGAGGAGAGCAACAGGCCCTCGATTATACAGTTCCAAAGTAAATTAAGTGAATTTGA 1021
Oy 3193 gctcaacagaagaatgacatgtgtccctgttgaagatlttgltgaatgtatcccaaggtgt 3252
Db 1020 ATTACACCGAAGATACATGTCTCCCATTTGGGAGATTTGGTGTAAATACCAAACTGCT 961
Oy 3253 tagccttgtattatgtgaatg-----ataaagatccaatgtatca 3292
Db 960 AGCCCTAGATTAATGAGAAACATGATGATGTAATCTGTATATACCAATTAATGATG 901
Oy 3293 aatgaacacagttccta-----gtcatltaaagcttgaactgtgccttaaaactag 3343
Db 900 AATGAACACTAGTTCCTTAATTAATTTATCTTTTAAAGCTTACCTGCTTAAACTAGA 841
|||||
```

```

OY 3344 gatcaattttctcgaactgagaaactttagccttccaacagttcacacctca----- 3397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 840 GATCAACTTTCACACTGCGAAAAGCTTTCAGCTTTCAGAAAGATCATCTTAAGAAAT 781
OY 3398 --gaaagtcagatattatttaccagactctcttggaacatctgcccccaatttaatt 3455
      | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 780 TGCACAGTAAGCATTTATTTTTCAGNCCATTTTGGACATCTCATCTTAATTAATTAAGT 721
OY 3456 ca-----tgtaggttagtattattacaa-aaaatgattgaatatagctgtcctt 3509
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 720 ATTCCCTGCTGCTTTAGTATTATTACAATTAAGGTTTGAATATATAGCTGTCTTT 661
OY 3510 atgcataaataaccagttgagcaactactgcccagagagaaaaagtttaagt-gctc 3568
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 660 ATGCATTAATAAACCCAGCTAGAGCATTACTGCCAGAGAAAAAATCGATTGAATGCGC 601
OY 3569 atttcctaccctcaaaagat-----aacgtaattattttggttaactaaagaatgcagt 3622
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 ATTTCCTTAAGTATTAGAGTCTCANATCGAATTAATTTGGCTATCACTAATAAATGCGT 541
OY 3623 atatttagtttccatttgcatagcgtgttggctatagacaatatttaattgaaa 3682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 540 AATATTAGTTTCCATTTCATGATGTTGTGTGCTATAGATATTTTAATTGAATA 481
OY 3683 attgttttaaatatttcttaccagtgaaagactgtttccaagctctttatatgtacata 3742
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 GTTTTGTAAATTAATTTTTCAGTGAAGACGTGTTTCAGCTCTTTTATTTGTAACATA 421
OY 3743 gactttttagtat---ctggcatatgtttgttagaccgttaagactgatatcttc 3799
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 GCTTTTATGTAATTTACTGCGCATATGTTTGTAACTGTTTAATGAAGACTGGA-TATCTTC 362
OY 3800 ctccaacttttgaatatcaaaaacagtgctttatact 3836
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CTTCAACTTTGAATFACAAAACAGTGTTTTACT 325

RESULT 4
US-60-172-360-693
: Sequence 693, Application US/60172360
: GENERAL INFORMATION:
: APPLICANT: MORRIS, MacDonald
: APPLICANT: Lal, Preeti
: APPLICANT: Dlep, Dinu
: TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
: FILE REFERENCE: GX-0007 P
: CURRENT APPLICATION NUMBER: US/60/172,360
: CURRENT FILING DATE: 1999-12-16
: NUMBER OF SEQ ID NOS: 29838
: SOFTWARE: PERL Program
: SEQ ID NO 693
: LENGTH: 3156
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 407511.3
: FEATURE:
: NAME/KEY: unsure
: LOCATION: 1563-1584
: OTHER INFORMATION: a, t, c, g, or other
US-60-172-360-693

```

	Query Match	42.4%	Score 1641.6;	DB 42;	Length 3156;
	Best Local Similarity	78.9%;	Firm. No. 1.8e-296;		
	Matches 2464; Conservative	0;	Mismatches 451;	Indels 209;	Gaps 35;
Oy	904	aagcattgttcatatctgagtattttagaacccaagaccattctcaagtgtgcga	963		
b	1	aagcattgttcatatctgaatttcacaaaagccaaagaccattctcaagtgtgcga	60		

QY	964	aggaataataatcccgagcaggttccagcgcgtctctgtgccaataatcagaattgttcag	1023
Db	61	aggaatatatactccggacaaatccagccactctctgcatacaaatatcaatgcctgttcag	120
QY	1024	ataagaagaggaataactcttcctgaattatctcaataataatagaatcccttggagcggttc	1083
Db	121	ataaggaagaggaataactcttcctgaattatctcaataataataagaatcccttggagcggttc	180
QY	1084	caggatattccaaaggatctccctcagttgtcatgtgtctcttgcacaagacattctgcagatt	1143
Db	181	cgggacattccaaaggatataatcagtgcatgtgtctcttgcacaagacattctgcagatt	240
QY	1144	gtaataataaagattgattgtggaagcgtcttcgtggaacattttatacagtgatctctc	1203
Db	241	gtaataataaagattgattgtggaagcgtcttcgtggaacattttatacagtgatctctc	300
QY	1204	gtgtccacagatgtcccgacgtgatgagccaattgcatactgaagccagagattgtctctc	1263
Db	301	gatgtccatggtgtcccgacgtgatgagccaagcgtgtctatactgaacacagagattgtcttc	360
QY	1264	ttgtgtgaaacttaaccgaacagttcatagagccatgaaglatgacaaagatgaagttg	1323
Db	361	ttgtgtgaaacttaaccgaacagttcatagagccatgaaglatgacaaagatgaagttg	420
QY	1324	acctcccatattgtattggatctctctctctgaaatgtgagaccagtgacactattccaatt	1383
Db	421	acctcccatattgtattggatctctctctctgaaatgtgagaccagtgacactattccaatt	480
QY	1384	ctataccccaatgaagtgccctcaaatataataaataagggacacctgtgcctatacatt	1443
Db	481	ccatcccccaatgaagtgccctcaagatataatataatagaacacctgtgcctatacatt	540
QY	1444	ttgatgtgagagctctctcgtggagactgcgagtgtaataatagattgtgtgataagcgctag	1503
Db	541	ttgatgtgagagctctctcgtggagactgcgagtgtaataatagattgtgtgataagcgctag	600
QY	1504	gtgtgtgaaatgaagccaacaccttgtgtgtataccctgtlaagaccttccaagaataatcagtga	1563
Db	601	gtgtgtgaaatgaagccaacaccttgtgtgtataccctgtlaagaccttccaagaataatcagtga	660
QY	1564	ctccacgcccacaagaagattgtgtcatattatcagagttgcccacaacaccttccata	1623
Db	661	ctccacgcccacaagaagattgtgtcatattatcagagttgcccacaacaccttccata	720
QY	1624	tttcggaagactcaagttccaccggaagaagctgtaccacaagaacctctctgtgattgtga	1683
Db	721	tttcggaagactcaagttccaccggaagaagctgtaccacaagaacctctctctgtgattgtga	780
QY	1684	caattgttagaccgaagccaacaacaagaatgttaatgatttagaaglatctgatatca--	1740
Db	781	caattgttagaccgaagccaacaacaagaatgttaatgatttagaaglatctgatatca--	837
QY	1741	gtttgtgtggaagaaanaaccacaagaagttcagactgatagtatgtgaagaaatttaag	1800
Db	838	gtttgtgtggaagaaanaaccacaagaagttcagactgatagtatgtgaagaaatttaag	897
QY	1801	-----tggaaaataccgaagatttlaaagcgtgtgtgttccagtagctgcgaacaaaatgaa	1854
Db	898	aacagatggaanaatccgagtttgtaagaatgtgttcttagactgtgggagagaaaaatgaa	957
QY	1855	gaacttcaattgtgcagaaacagtgatgaanaatgtctgcgctataatgactgtcagaagcgaga	1914
Db	958	gaacttcaattgtgcagaaacagtgatgaanaatgtctgcgctataatgactgtcagaagcgaga	1017
QY	1915	tttagtaagcggctgtgaagtgaaatcaatactcgtgttgacccaataatcgtataatctc	1974
Db	1018	tttagtaagcggctgtgaagtgaaatcaatactcgtgttgacccaataatcgtataatctc	1077
QY	1975	acggatgctgaggtatatactcagactctgaagatgaagcgtctgtctctcagttccctgtgcga	2034
Db	1078	atcgccgtgaggtatatactcagactctgaagatgaagcgtctatacctctcagttccctgtgcga	1137

OY	2035	gtaacagtgacagtgagacacgtgcgaagttccaagttta- gaagaaccccttggaagatgaa	2035
Db	1138	gtaacagtgaaagtgaggacatgcgcagaagttccaagtttaaggaagaacccatggaagttaa	1197
OY	2094	agtcataattgaagaattctcaaatgctctggaaagtatcagcggagaagcccgaaatgctc	2153
Db	1198	agtgaaattgaaagaattctcaaatgctctagaagaatgagcccgatgctccagagaagact	1257
OY	2154	ggaagatctcgattcttgagccttgatggaagggaatcgaagaagtttgttaatgaaagctaatg	2213
Db	1258	ggaagagctcgattcttgagccttgatggaatgatacgaaggcaattaatgaagctatatct	1317
OY	2214	acaagaacagaattgtgaagaagtgtgaacatcatccatcgaacaaatcataactattgaagc	2273
Db	1318	gtgaacaacagaagatgaacaagacatgaacatcatccatacaaaatcat-- agtgaataat	1374
OY	2274	gtcccgagatccaagaattgtgtccacagcacttgaggaaacttgatgcacgtgcacaaaatgaa	2333
Db	1375	gtgtcaggtacagaatgtgttccaacagcatctagaagaacttgacgtgc-- aaaaatgaa	1431
OY	2334	tgattactctgtaac---ctgaacaaggaaatctgaaagaatgataattatagacgtg	2389
Db	1432	tgattactctgtaacactcgatagagcaagaacacagaaggtgaataattatagtggtg	1491
OY	2380	-aaatatagatg- tctctctgataatcttcaaatgtccatcaatctctgttgtaactgtg	2447
Db	1492	taaaatagatgtgttttccaagataattttaactctatattctctactgttacaac	1551
OY	2448	acattcaaacctgtgtgtgact-----	2470
Db	1552	tcaacacaaacnnnnnnnnnnnnnnnnnnnnngtactaagtatcttcaatcagctgt	1611
OY	2471	-----tcatctctcttccaagttcatgtgatactgataactcgatgtatgata-	2520
Db	1612	gtccaagactaacctctctttaaagttcaatgtatctgatactgaataatcattgtatataat	1671
OY	2521	---atttgttttttgccataatgagttccaaccttt-- aagtttccaaaagccatgtg	2575
Db	1672	aatttttttgtttgtctagtgagttccaacctttttaagcttttccaaaagccatgtg	1731
OY	2576	aatg-----ttaatgtaagaaggaaagcgtatcatctgacccaagaatgttatccaact	2630
Db	1732	aagttcaaataaatgaataaggaaagcagcttaactctagaaccaagaatgttatct--tca	1788
OY	2631	tttgtttgttaacaatgaatgaattttaaagccccaatcttctctctgcgaacttta	2690
Db	1789	ttctcttctgaacacttgaatggttttgaaagtactcaaaatctgttaacgttaaaactttgat	1848
OY	2691	tttagacagtttaactttttaaacactgycathtccaanaactgttgacgactaacttt	2750
Db	1849	tctttaacaacatatatttttaaacactgycathtccaanaac- tgtggcagctaaacttt	1907
OY	2751	taaaatcaacgatgacttgttaatgtga-----ggagtcagcacccgtctgagacact	2803
Db	1908	taaaatctcaaatgacatgacgtgtgtagtaagaagaatcacaataatgttggaagagca	1967
OY	2804	c-----	2824
Db	1968	ctcggtgtccttctctttaaagaataactgtgtgctagaataatccagattatgtat	2027
OY	2825	gaagcgtactactgacatgctttctgtactgtctgagac-- gtgtaatagtccaacag	2882
Db	2028	ttacgttcaaatgaaatgagctgtctgtactcctgtgagacatgtagtaatgtcataatg	2087
OY	2883	ggcccttgaaactaatctgataaagtatt-----tgaataatgtgttcaagtgtctaga	2936
Db	2088	gtcataaaacataacacgaaaacaaataaagcttttgaaatgtttcaagtgtcttga	2147
OY	2937	aaacaatgtgcctgtcctataatgagtcocccatgattggaatatttgccattgttaataat	2996
Db	2148	aaacattgtgcctgtccgtgac---cccttaagtttggaataatttgcgcattgtgttttaa	2204
OY	2997	atacctatcaactgtgtgtagagcctgtgataagatcttccacacaaat---actgcgaaga	3051

[illegible]

RESULT 5
US-09-496-914A-4224
: Sequence 4224, Application US/09496914A

```

; GENERAL INFORMATION:
;
; APPLICANT:      Tang, Yuanhua T.
;
; APPLICANT:      Tillinghast, John
;
; APPLICANT:      Sinku, Ankura
;
; APPLICANT:      Liu, Chenghua
;
; APPLICANT:      Drmanac, Radotje T.

```


;; TITLE OF INVENTION: Novel Configs Obtained
;; FILE OF INVENTION: From Various Libraries
;; FILE REFERENCE: 787
;; CURRENT APPLICATION NUMBER: US/09/496,914A
;; CURRENT FILING DATE: 2000-02-03
;; PRIOR APPLICATION NUMBER: US 09/353,690
;; PRIOR FILING DATE: 1999-07-14
;; PRIOR APPLICATION NUMBER: US 09/034,341
;; PRIOR FILING DATE: 1998-02-13
;; PRIOR APPLICATION NUMBER: US 09/045,400
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: US 09/321,214
;; PRIOR FILING DATE: 1999-05-26
;; PRIOR APPLICATION NUMBER: US 09/131,598
;; PRIOR FILING DATE: 1998-08-10
;; PRIOR APPLICATION NUMBER: US 09/431,517
;; PRIOR FILING DATE: 1999-11-01
;; PRIOR APPLICATION NUMBER: US 09/328,351
;; PRIOR FILING DATE: 1999-06-04
;; PRIOR APPLICATION NUMBER: US 09/332,782
;; PRIOR FILING DATE: 1999-06-14
;; PRIOR APPLICATION NUMBER: US 09/235,076
;; PRIOR FILING DATE: 1999-01-20
;; PRIOR APPLICATION NUMBER: US 09/234,611
;; PRIOR FILING DATE: 1999-01-22
;; PRIOR APPLICATION NUMBER: US 09/346,956
;; PRIOR FILING DATE: 1999-07-02
;; PRIOR APPLICATION NUMBER: US 09/362,510
;; PRIOR FILING DATE: 1999-07-27
;; PRIOR APPLICATION NUMBER: US 09/240,371
;; PRIOR FILING DATE: 1999-01-29
;; PRIOR APPLICATION NUMBER: US 09/248,797
;; PRIOR FILING DATE: 1999-02-12
;; PRIOR APPLICATION NUMBER: US 09/271,490
;; PRIOR FILING DATE: 1999-03-18
;; PRIOR APPLICATION NUMBER: US 09/293,972
;; PRIOR FILING DATE: 1999-04-15
;; PRIOR APPLICATION NUMBER: US 09/274,861
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: US 60/125,453
;; PRIOR FILING DATE: 1999-03-19
;; PRIOR APPLICATION NUMBER: US 60/126,605
;; PRIOR FILING DATE: 1999-03-26
;; PRIOR APPLICATION NUMBER: US 09/306,350
;; PRIOR FILING DATE: 1999-05-07
;; PRIOR APPLICATION NUMBER: US 09/399,720
;; PRIOR FILING DATE: 1999-09-21
;; PRIOR APPLICATION NUMBER: US 09/404,284
;; PRIOR FILING DATE: 1999-09-21
;; PRIOR APPLICATION NUMBER: US 09/465,877
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
;; PRIOR FILING DATE: 1999-01-19
;; NUMBER OF SEQ ID NOS: 10410
;; SOFTWARE: PC_CT_genes Version 1.02
;; SEQ ID NO 4224
;; LENGTH: 1667
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(1667)
;; OTHER INFORMATION: similar to g1525318 in the genepept database release 114,
;; OTHER INFORMATION: Run with FASTX 3.3f00, default parameters
US-09-496-914A-4224

Query Match 33.9%; Score 1312; DB 18; Length 1667;
Best Local Similarity 88.7%; Pred. No. 5.9e-235;
Matches 1481; Conservative 0; Mismatches 175; Indels 14; Gaps 5;
QY 600 atgattggaccagatcctcgacaattcttaagattattaccagaagaattcctcca 659
|||||

Db 1 atgattggcacagatcctcgacaattccttaagaattattatgcccggaaacaatacctcca 60
QY cctgagcctgagatgatacagctgtggcagatttatttaataatccttcgaaccacca 719
|||||
Db 61 cctgagcttgatgatatgacactgtggcagatttatttaataatccttcgaaccacca 120
QY 720 aagcggaaaaaaagaagatalcaatacaattgaagatgtgtgaagtcttcgacagag 779
|||||
Db 121 aaaggaaaaaaagaagaatattacaattgaagatgctggaaattactgcaagag 180
QY 780 tgaataaagaataagttctgactggagctgggtttctgctcctgtggattcccgac 839
|||||
Db 181 tgcataaaaattatagttcttaactggagctgggtgtgtctgttcaatgtgaaattctgac 240
QY 840 ttcaagtaagaagcagatcatgctcgccttcgggtggagactccgaacctccagac 899
|||||
Db 241 ttcaagtaagaagcagatcatgctcgccttcgggtggagactccgaacctccagac 300
QY 900 cctcaagcagatgttgaatattgaatatttagaagaaccacaacattcctcaagttc 959
|||||
Db 301 cctcaagcagatgttgaatattgaatatttagaagaaccacaacattcctcaagttc 360
QY 960 gcaaaagaaataatcccgagcagttccagccgctcctgtgtccaaattcatagcttgc 1019
Db 361 gcaaaagaaataatcccgagcagttccagccgctcctgtgtccaaattcatagcttgc 420
QY 1020 tcagataaaggaagaaactccttcgaatataactcaaaatagatacttggagcag 1079
|||||
Db 421 tcaagataaaggaagaaactccttcgaatataactcaaaatagatacttggagcag 480
QY 1080 gttgcaggaatccaaagatccttcagatgttcaatgttcccttgaacagcattctgcctg 1139
|||||
Db 481 gttgcgggaatccaaagataatccagatgttcccttgaacagcattctgcctg 540
QY 1140 attgtaataatcaagatgttgaatgttgcgtgtggaagcattttaaacaagtgtc 1139
Db 541 attgtaataatcaagatgttgaatgttgcgtgtggaagcattttaaacaagtgtc 600
QY 1200 cctcggtgccttagtggtgccagcctgagatgagccacttgcacatgaaagcagagatgtc 1259
|||||
Db 601 cctcgatgtccttagtggtgccagcctgagatgagccacttgcacatgaaagcagagatgtc 660
QY 1260 ttcttgggtgaaaacttccagaaagtttcatagagcagatgaaatgaaatgaa 1319
|||||
Db 661 ttcttgggtgaaaacttccagaaagtttcatagagcagatgaaatgaaatgaa 720
QY 1320 gttgacctcccatgttattgtgattcttcttgaaagtgaagcagatgacctaatcca 1379
Db 721 gttgacctcccatgttattgtgattcttcttgaaagtgaagcagatgacctaatcca 780
QY 1380 agttctatacccatgaaagtgtccccaattataataagaggaaccttgcctcatcta 1439
|||||
Db 781 agttccatacccatgaaagtgtccccaattataataagaggaaccttgcctcatcta 840
QY 1440 cattttgatgtagagctccttggagactgcagtgttcaatlaattgagttgtgcatag 1499
|||||
Db 841 cattttgatgtagagctccttggagactgcagtgttcaatlaattgagttgtgcatag 900
QY 1500 ctgagtggtgataatgccaaccttgtgttaacctgtaaagccttcgaagaataacgaa 1559
Db 901 ttgagtggtgataatgccaaccttgtgttaacctgtaaagccttcgaagaataacgaa 960
QY 1560 aaactccagcccaaaaagaattgttcaattatcaagattgcccacaaaccttc 1619
|||||
Db 961 aaactccagcccaaaaagaattgttcaattatcaagattgcccacaaaccttc 1020
QY 1620 cataattcgaagaactcaagttcaccttgaagaactgtaacacaagaactctctgtgatt 1679
|||||
Db 1021 catgtttcagaagactcaagttcaccttgaagaactgtaacacaagaactctctgtgatt 1080
QY 1680 gtaacaactgttagacccaagcaacaacaatgttaattagaaagtatcgaatca 1739
|||||
Db 1081 gtcaacactttagacccaagcagctaagaagtaatg---atgatttagatgttctgaatca 1137

[illegible]

Query Match	29.0%	Score 1122.8;	DB 17;	Length 2608;
Best Local Similarity	75.3%;	Pred. No. 1.6e-199;		
Matches 1926;	Conservative	0;	Mismatches 422;	Indels 210; Gaps 34;
QY 1469	cgatgttataataatagtgctgttcataagctcaggtgtgtgtgaatagccaacttgttg	1528		
Db 18	ctatgacacaaatcctaatatgaatgacataggttaggtgtgtgatatgccaacttgcgt	77		
QY 1529	taacctgtaaagcttcacgaataatcagaaaaacctcacgcccccaaaaagaattggt	1588		
Db 78	--acctgtaaagcttcacgaataatcctaaacacctccacgaacaaacaaaagaattgac	135		
QY 1569	tcattatcaggtgtgcccacaaacctcttatatttcggagaagactaaagtccctga	1648		

Db	136	ttatttgcagagtgctgcaccacacaccttcaatgttttaagaagaactcaagttaccaga	195
Qy	1649	aagaactcgtacacaaagaactctctgtatgtgtcaactctgtagaccagacacaacaa	1708
Db	196	aagaactcaaccacagattcttcagttgattgtcacacttttagaccagacgtctaag	255
Qy	1709	caatgcttaatgatttaagaatcgtaaatca--gtctgtgtgaagaanaaacacaaga	1765
Db	256	taatg---atgatctaaatgtgtcgtatcaataaaggctgtatgtgaagaanaaacacaaga	312
Qy	1766	agtcacagactcgtaggaattgttagaacttaatg-----tggaaaaatccagatttta	1819
Db	313	agtcacaactctcgtaggaatgtgaagatcattgtcgtacaacagatgtgaataccgattgtaa	372
Qy	1820	ggtctgtgtgtccagatctgcagacacaataatgaagaagaactcagctgtgaagaacgttag	1879
Db	373	gaatgtgtgtctcgtactcgtgtgggaagaataatgaagaacttcagtgctgtgaacgttag	432
Qy	1880	aaaatcgtgccttaatgactctgcagaagagacagattagtaagcgtcttgaggtaatca	1939
Db	433	aaaatcgtgccttaatgactctgcagaagagacagattagtaagcgtctgtagtaatca	492
Qy	1940	atactgtttgttaaccaccaaatctgttacaattccacggtgtctgtgtgtatactcaagtc	1999
Db	493	gtactgttttgcaccacaatactgttaacattccctcctgtgtgtgtgtatattccagacc	552
Qy	2000	tgaagatgaactctgtcgcctcattccgttccgtgtgcataaagtgacaatgtgcacatgca	2059
Db	553	tgaagatgaactctcattccctcctcgtctcgtgtgcataaagtgatagttggaacatgcca	612
Qy	2060	gagtcacaagtttaagaagaacctctgtgaagaatgtgaagtgaaatgtgaagatttacaacgtg	2119
Db	613	gagtcacaagtttaagaagaacctctgtgaagaatgtgaagtgaaatgtgaagatttacaacgtg	672
Qy	2120	cttgcgaagatgatacgcgaagagcgcgcgaatgtcgtgtgaggaactcgtatcttgagctgtatg	2179
Db	673	cttgcgaagatgatacgcctgtatgttccacaagagagcgtgtgagagctgtgattcttgagagctgtatg	732
Qy	2180	aggagatcaagaagttgttgaatgaactgaagctgaagcagacggagattgcacagatgtaa	2239
Db	733	aggatcaagaagagcgaatttaatgaagctatactcgtgtgaacacggagaagtacaacagatagtaa	792
Qy	2240	ctatccatcagacacaatcaataactatitgaagctgtccggaattcagaatgtgtccacc	2299
Db	793	ctatccatcaacaacaatcat--agtgtaatattgtgtgaggtacaaggaattgtgtccacc	849
Qy	2300	agcatcttggaactctagcatgttcaaaaaatgtaatgttactcttgtgaac----ttgaaca	2355
Db	850	agcatcttggaactcttagcatgtc--aaatgtatgttctactctgtgaactcgtatgtgaaca	906
Qy	2356	aggaaactcgtgaagaatgtatatattatagaactgtgaaatagaatgtg--ctctctgtgaata	2413
Db	907	aggaaacccagaaggtgtataacttctgaagttgtgtgaataaagatcgtttcttcaatgtgaata	966
Qy	2414	ttcttaagttccatcttctgtctgtgtactgtgtatcaatcaacactgtgtgtgtact--	2470
Db	967	tttttaactcatattattctgtactcgttaacaaactcaacaactatcttttttttaa	1026
Qy	2471	-----tcaatctcattcaag	2486
Db	1027	aaaaaaaagttacaaagtatccttcaactcagcgtgtgtgaagaagaactccttctttaaag	1086
Qy	2487	gttaattgtatgttaactcgtcgtatgtatgata---atttgttttcttctaataag	2542
Db	1087	gttaattgtatgttaactcgtcgtatgtatgataatcttttctgtctgtcgtatgag	1146
Qy	2543	tttcaacctttaaagttt--caaaagccatgtgaatg----ctaaagttaaggagac	2593
Db	1147	tttcaaccttttaagtttcaaaaagccatcgtgaatgttaaatatgttaaggagac	1206
Qy	2556	agcttactcagacaagaatgttatctcaacacttttgttgtgttaacatgtaaatgtt	2655
Db	1207	agcttactcagacaagaatgttatctt--tcaacttcttctgttaacatgtaaatgtt	1263

QY 2656 aaagccctcaattctgtctcgtgaacttatttttagagacagttaacttttaaca 2715
 1264 gaagtactcaaaatctgttaagcctaaactttgttcttcaacaaattattttaaca 1323
 QY 2716 ctggcatttccaaaactgtgacagtaacttttaaatcacagatgactgttaagt 2775
 1324 ctggcatttccaaaac-tgtgacagtaacttttaaatctcaaatgacatgctgt 1382
 QY 2776 ga-----ggagtcagcacgcgtgtctggagcactcaaaactgtgctcagtggtgaag 2828
 1383 gaggagaaggaagcacaacatactggagagcagccggtgtcttactttaaagtc 1442
 QY 2829 cgtacttactgc-----atcgtttt 2849
 1443 aatactgtgtcgaagaattcagattatgtattttagtcaatgaagaatgtctttt 1502
 QY 2850 gtaactgtctgacagc--gtgttaattccaaacagccctgagacttaactgtataatg 2907
 1503 gtaactcctgtgacagctgttaattgtctatgtctcctaaactaaactgtaaaaaa 1562
 QY 2908 att-----tggaaatgtttcagttgtcttagaacaacatagtcgtctctataggt 2961
 1563 aataaatgcttggaaatgttccagttcagtttagaacaacatagtcgtcctgagat---c 1619
 QY 2962 ccccttagttgaataatttgcacattgtttaataatacctatcactgtgtgtagagctg 3021
 1620 ccccttagttgaataatttgcacattgtttaataatacctatcactgtgtgtagagctg 1679
 QY 3022 catagatcttccacaacaat-----actgcaagaatgtgaataatgcaaaagccttcgaa 3076
 1680 catgactcttccacaagaatgttaactgccaagaatgtgaataatgcaaaagccttcgaa 1739
 QY 3077 tc--taataatgtgactctcactgaggagagtgtaataattttagacgtcttttccat 3144
 1740 tctataaatgtgactctcactgaggagagtgtaataattttagacgtcttttccat 1798
 QY 3135 taatgaagaaagcaatagactcttaatt--taaatgccaagtcataaga--taaatgt 3191
 1799 taatgaagaaagcaatagactcttaatt--taaatgccaagtcataaga--taaatgt 1858
 QY 3192 agctcaacccgaagaatgacactgtgtcgtgtgagagattgtgttaatgtcccaagtg 3251
 1859 aatcagcccggaagaatgacactgtgtcgtgtgagagattgtgttaatgtcccaagtg 1918
 QY 3252 ttagcctgtatattgagagatga-----atagacatcccaatagtc 3291
 1919 tagccctagatttatgtgagatgacatgagatgttaactgttaatgacgaatagttcaat 1978
 QY 3292 aaatgaactagttctta-----gttaattaaagctttagctgtcttaaaactag 3342
 1979 gaatgaactagttcttaattatcttaatttaaaagctttagcctgtcttaaaactag 2038
 QY 3343 ggaatcaatttctcaactgtcagaacttttagccttcaaaagttcacacactcaga-- 3400
 2039 agatcaacttctcaactgtcagaacttttagccttcaaaagttcacacactcaga-- 2098
 QY 3401 -----agtcagtaattatttacaagacttcttggaacattgcccacaatttaaatat 3454
 2099 ttgcacagtaagcattatatttcaagacattttagaacaactcactcaatataaaga 2158
 QY 3455 t-----catgtgggtttagattattatataca--aaaaatgtttgaaatagctgttctt 3508
 2159 tatctccctgttgcatttagattattatatacaataaaagggtttgaaatagctgttctt 2218
 QY 3509 tatgcaataaataaccagtttagagcaatctcgtccagaggaagaatataatga--gct 3567
 2219 tatgcaataaataaccagtttagagcaatctcgtccagaggaagaatataatga--gct 2278
 QY 3568 catctccctccctaaagaat-----aactgaattatttggcactaaagaatgag 3621
 2279 catctccctccctaaagaatctcaatctgaattatttggcactaaagaatgag 2338

QY 3622 tataattagtttccacttgcatagtgtgtgtgtcctatagacataatttaaatgaaa 3681
 2339 tataattagtttccacttgcatagtgtgtgtgtcctatagacataatttaaatgaaa 2398
 QY 3682 aattgttttaaatatttattttagagtgaaagactgtttttagcctttttatatgtcat 3741
 2399 agttgttttaaatatttattttagagtgaaagactgtttttagcctttttatatgtcat 2458
 QY 3742 agacttttagtaat-----ctggacatagtttttagagccgtttaaagtcagtaactct 3798
 2459 agcttttagtaatattttagagtcagatagtttttagagcgttttaagtcagta--tact 2517
 QY 3799 cctccaaacttttgaataacaaaacagtgtttatact 3836
 2518 cctccaaacttttgaataacaaaacagtgtttatact 2555

RESULT 7
 US-09-315-788-2235
 ; Sequence 2235, Application US/09315788
 ; GENERAL INFORMATION:
 ; APPLICANT: Gearling, David P.
 ; TITLE OF INVENTION: NOCULEIC ACID MOLECULES DERIVED FROM A
 ; FILE REFERENCE: MLN98-19PM
 ; CURRENT APPLICATION NUMBER: US/09/315,788
 ; EARLIER FILING DATE: 1999-05-21
 ; EARLIER APPLICATION NUMBER: 60/086,455
 ; EARLIER FILING DATE: 1998-05-22
 ; EARLIER APPLICATION NUMBER: 60/132,067
 ; EARLIER FILING DATE: 1999-04-30
 ; NUMBER OF SEQ ID NOS: 2346
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2235
 ; LENGTH: 2678
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(2678)
 ; OTHER INFORMATION: n - A,T,C OR G
 US-09-315-788-2235

Query Match 28.7%; Score 1111.8; DB 17; Length 2678;
 Best Local Similarity 75.38; Pred. NO.1.8e-197;
 Matches 1926; Conservative 0; Mismatches 422; Indels 211; Gaps 35;

QY 1469 cgaatgataataatgagttgtgtcactagcgtggtggaatgcaaaactgtgtg 1528
 20 ctatgaacaactatcatatattatgacatagtgtagtgggtgaatgcaaaactgtgtg 79
 QY 1529 taacctgttaagcttccagaataattactgaaaacccctcagcccaaaaagaatgtgt 1588
 80 -accctgttaagcttccagaataattactgaaaacccctcagcccaaaaagaatgtgtg 137
 QY 1589 tcattttagagagttgcccacaacactcttcataatttggaaagtcacagttcaactgtg 1648
 138 ttattgtcagagttgcccacaacactcttcataatttggaaagtcacagttcaactgtg 197
 QY 1649 aagaactgtacacaagaactcttctgtgattgtctacactgttagacccaagaacaacaa 1708
 198 aagaactgtacacaagaactcttctgtgattgtctacactgttagacccaagaactgaag 257
 QY 1709 caatgttaatgatttagaagatctgaaatca--gttgtgtggaagaaacccaaga 1765
 258 taatgt--atgatttagatgtgtctgaaatcaaaaggtgtgataagaacaaacacaga 314
 QY 1766 agtaacagtagtagaagattttagaacaattatgt-----tggaaatccagattttaa 1819
 315 agtaacagtagtagaagattttagaacaattatgtctgtgaacagatggaataatccgatttga 374

OY	1820	ggcgcgttgcgttcacgtactgcgcagacaacaatgaaagaacttcagttgcagaaacagtgag	1879
Db	375		434
OY	1880	aaatgcgcgcctaaatagactctgcaagaagacagatctaaagccgcttgagggttaaca	1939
Db	435		494
OY	1940	ataccgtttcttcaaccaacaactcgtttacatattccacgcgttgcttgaggatatactcagactc	1999
Db	495	gtatcgttttttgcacccaataatcgttaacatttccatcggcgtctgggtatatactcagactc	554
OY	2000	tgaagatgcgcgtctctgtcctctagcttcctcgttgccgttgagtaacagtgacgtgagacatgcga	2059
Db	555	tgaagatgcgcgtctctcctctagcttctctcgttgccgttgagtaagtgagacatgcga	614
OY	2060	gaatcccaagttttgaaagaagaacctcttgagaagatgaaagtgaatt--gaagaattctacaagt	2118
Db	615	gaatcccaagttttgaaagaagaacctcttgagaagatgaaagtgaattctgaaatctctacaagt	674
OY	2119	gcttcggaagaatgcgaacggaagagcccggaatgcgtctggaagacatcttgatttgagatgtgag	2178
Db	675	gcttcagaagaatgcgaacctgcgatgtctccagagagacgcgagagagcgtggaattcgagacctgcgtg	734
OY	2179	gagggagatcacaagaagtgctgtttaatgaaagcctatagctctacaagacagagaatgcagagtta	2238
Db	735	gagatgctatcaagaagagcgaattaatgatgagcctatattctgtgaaacaggaagtacaagacatga	794
OY	2239	actatccatcagacaataatcaataacactatgaaagctgcgtcccggaattccagaagtgtctcac	2298
Db	795	actatccatcacaacaataatcat---agtgtaaatattgtgcagatccaggaattgtctcac	851
OY	2299	cagacttcggggaacttttgacatgctcaaaaaaataagatggttacttgagac---ttgagac	2354
Db	852	cagacttaagaagaactttcagctgtcc---aaaaagaaagtttactcgtgcgaactgcgaataagagc	908
OY	2355	aaggaaatctgaaagaatgtaattatctatagactgcg--aaaatgaaatg--tctctctgagta	2412
Db	909	aaggaaacacgaagaagtgtaatatattatagtgtrtgtaaaatagattgtttcttcagtgta	968
OY	2413	attcttaaatgtccatcaattctcgtttgttaactgtgcatactgacaacacgtgttgtaact--	2470
Db	969	attcttaactcatattattcttgactgtgatacaaacctcaacataacttttttttttaa	1028
OY	2471	-----tcaatctcccttcaa	2485
Db	1029	aaaaaaaaaaggtactaagaatctcttcaatcagcgtgtgtgtcgaagactaaacttcttttaa	1088
OY	2486	ggtcattctgtatgatacactgcgtatgtaatgtaata---atttgtttttgcctaagta	2541
Db	1089	ggtcattctgtatgataatataatcatatgctgataataataatttttctgtttgtctcagta	1148
OY	2542	gttccaacctttaagtttt--caaaagccattggaatg----ttaatgtaaaaggaa	2594
Db	1149	gttccaacattttaagaatttttcaaaaaagccatcggaaatgtaaaattaatgtaaaaggaa	1208
OY	2595	cagcttctctagaccagaagaatgcttttcaacacctttttgtttgtaacatgtaagt	2654
Db	1209	cagcttaactctagaccagaagaatgtaattt---tcaacttctctgttaacatgtaagtgt	1265
OY	2655	taaaagccctcaaatctcgtctcgtctgaacacttatttttagagacggttaaacattttaaac	2714
Db	1266	tgaagtaactcaaaatctcgttagcgttaaacctttgtatctttaaacaacaattattttttaaac	1325
OY	2715	actggtcatttctcaaaaactctgtgagcagctaaactttttaaataatcacagatgtaagt	2774
Db	1326	actggtcatttctcaaaaac--tgtgagcagctaaactttttaaataatcacaaatgtaagt	1384
OY	2775	tga-----ggaatgagacacgctgtctgagagaaactcaaaacttggcgtcagtggtgaa	2827
Db	1385	tgaatgtaagaagaagtcttaacaataatgtgtggaagagcactcgtgtgtccttaactttaaag	1444
OY	2828	gcgtactactgc-----atcgtttt	2848

Db	1445	taatactctgctgcgaagaattcagagattatgtatattacgttcacaaatgaagatgcttt	1504
Qy	2849	tgtaactctgtcgaac--ggtgtaattgtccaaaacagcccgagacaaactgtataat	2906
Db	1505	tgtaactctctgtgcagctgaagaaagtcctataattggtccataaacaactcgaac	1564
Qy	2907	gatt-----tggaaatgctgttcgaattgtctcagaaaacatagcgctgtctataagg	2960
Db	1565	aaataaagctttggaatagtcttccagtgtgtttagaaacaattagcgctgcctgac--	1621
Qy	2961	tcoccttagtttgaataatgtgcattgttbaataaataccatacactgtgtgtagagct	3020
Db	1622	ccoccttagtttgaataattgtccattgtgttttaataaccatacactcgtgtgtagagct	1681
Qy	3021	gcatagattcttcaccacaat-----actgcgaagtgtggaatagcaaaagcctttcga	3075
Db	1682	gaattgactcttccacaagatataaactgtccaaaatgtgaataatgcaaaagcctttcga	1741
Qy	3076	atc--taataatgttactctcactcagggagagtgataattttggacgcgtttttcca	3133
Db	1742	atctataataatgttactctcactcagggagagtgataattttggacgcgtttttcca	1800
Qy	3134	ttaatgagaaagaataatgacgccttcaat--taaaagcccaaaagtcataaag--taaat	3190
Db	1801	ttaatgagagagagaacagccctgattatacagttccaaagaataaagatgttaattg	1860
Qy	3191	tagctcaacagaaagatcacactgtgtgcgtgtgaggaattgtgttaattgtcccaagt	3250
Db	1861	taattcagccagaaagatcacactgtgtgcgtgtgaggaattgtgttaataaccaaactg	1920
Qy	3251	gttagcctgtatattaggaatga-----atacagttccaatagt	3290
Db	1921	ctagcccttagatattatgagatgagatgacaatgatactgttaattagcgaatagttaa	1980
Qy	3291	caaatgaacactagttccta-----gttatttaaaagcttagctgtgccttaaaacta	3341
Db	1981	tgaatgaacactagttcctataattatcttatttaaaagcttagcctgtgccttaaaacta	2040
Qy	3342	ggagatcaatttcccaacgtcgagaacatttagcctttcaaaagttcaaccctcagaa-	3400
Db	2041	gagatcaacttcccaacgtcgagaacatttagcctttcaaaagttcaacccttattgaa	2100
Qy	3401	-----agtcagattattatttaacagactctctgtgaacatgtgcctccaatttaata	3453
Db	2101	attgcacagtaagaattatttttccagacatttccgaacacttccgaacactccaatttaata	2160
Qy	3454	tt-----catgtggttttagtatttatttaaca--aaaatgatttgaatatagcgttct	3507
Db	2161	gtattccctcgttcttagtatttatttaacaataaaaaggtttgaaatatagcgttct	2220
Qy	3508	ttaatgataaaaaataccagtttaggaacattactgtccagaggaagaatbatlaagta-gc	3566
Db	2221	ttaatgataaaaaataccagtttaggaacattactgtccagaggaagaatbatlaagta-gc	2280
Qy	3567	tcattccctcactcaaaaagat-----aactgattattttgtgcacactaaagaatgca	3620
Db	2281	ccattccctcactcaataaagatgtctccaacttgaattattttgtgcacactaaagaatgca	2340
Qy	3621	gtataattagtttccactttgcatgagtgttctgtcctatagacaatattttaaattgaa	3680
Db	2341	gtataattagtttccactttgcatgagtgttctgtcctatagacaatattttaaattgaa	2400
Qy	3681	aaattgttctaataatttttcaacgttgaagactgttttcagccttttatatgttaca	3740
Db	2401	aaattgttcttaataatttttcaacgttgaagactgttttcagccttttatatgttaca	2460
Qy	3741	tagacttttatgaat--t-tggacataatgtttgtagacggttttaagaactgtattact	3797
Db	2461	tagacttttatgaat--t-tggacataatgtttgtagacggttttaagaactgtattact	2519
Qy	3798	tcctccaacttctgaatatcaaaaacagtgttttataact	3836

Db	382	ACGTCAGAGGTGAG-----TATTTTACAGACTCTTTGGACACTTGGCCCCCAAAATTTAA	327
QY	3452	tatcatctgtygggttagtatttatacaaaaaaagatttgaatatagctgtctcttat	3511
Db	326	TATTCACGTTGGTTTACTATTATTAACAAAAAATGATTTGAAATATACCTGTCTTAT	267
QY	3512	gcataaaatcccaagtttaggacattactgcccagaggaagaaattattaagtagctcat	3573
Db	266	GCATAAAAATACCCAGCTAGAGCCGTTACACCCAGAGGA-TAAAGTATTAAGTACCTCAAT	208
QY	3572	tccttacctataaagat-----aacgtatattttgtgcctacctataaagaatgcaatat	3626
Db	207	TCCCTACCTAAATATCTCAATCTGAATTTATTTGGCTACACTAA-NATSCACTATAT	149
QY	3627	ttagtttccatttgcatgtagtgttttgctatagaacaattttaattgaanaattt	3686
Db	148	TTATGTTTCCATTTGCAATGATGTTGTGCTATTAACAATATTTTAAATTTAAAAAATTT	89
QY	3687	gtcttaaatatttttacaatgaagaagctgttcacgctctttatattgtacataagct	3746
Db	88	GTTTAAATATATTTTTCACGTGAAGACTGTTTTTCACCTCTTTTATATGTACATAGACT	29
QY	3747	tttatgtaactcgtgcatagtttttagag	3774
Db	28	TTTATGTAAATCTGGCATAATGTTTTTAGAG	1

```

RESULT 13
US-09-649-162-4406
; Sequence 4406, Application US/09649162
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: Richardson, Jennifer
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1181-001
; CURRENT APPLICATION NUMBER: US/09/649,162
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,057
; PRIOR FILING DATE: 1999-08-27
; NUMBER OF SEQ ID NOS: 9990
; SOFTWARE: FastSeq for Windows.Version 4.0
; SEQ ID NO 4406
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(503)
; OTHER INFORMATION: n = A,T,C or G
; US-09-649-162-4406

```

Query Match	10.1%	Score 390.2	DB22	Length 503
Best Local Similarity	91.2%	Pred. No. 8.1e-63		
Matches 436	Conservative	0	Mismatches 39	Indels 3
				Gaps
QY 1129	catctgcctgattgtgaataacaagattgattgtaagctgttcgttgagacatttta	1188		
Db 18	cgcttcgcctgattgtgaataacaagattgactgtgagactgtaagcagatatttta	77		
QY 1189	atcaggtagttcctcgtgtgcacctagttgccagctgaatgaagcaattgcatcatgaagc	1246		
Db 78	atcaggtagttcctcgtgtgcacctagttgccagctgaatgaagcaattgcatcatgaagc	137		
QY 1249	cagagattgtcttccttcttggtggaacaacttacaagaacagttcatagagcatgaaatg	1306		
Db 138	cagagattgtgttttttggtaaaattttaccagaaacagtttcatagagcatgaaatg	197		
QY 1309	acaaagatgaagtgtgacctccatcatgttatatggaattctctctgaaagtgaagccaagt	1366		
Db 198	acaaagatgaagtgtgacctccatcatgttatatggaattctctcccttcaaaatgaagccaagt	257		

PT New peptides useful for diagnosis, prevention and treatment of cancer
PT and immune disorders -
XX
PS Claim 1; Page 120-121; 193pp; English.

CC 7Y3325-13389 are human transcriptional regulator molecule (HTRM) protein
CC sequences. The HTRM protein and nucleotide sequences are useful for
CC preventing or treating disorders associated with decreased expression or
CC activity of HTRM which include cell proliferative disorders such as
CC actinofibrosclerosis and cirrhosis; cancers including adenocarcinoma and
CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes
CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists
CC of the HTRM polypeptides are useful for treating or preventing disorders
CC associated with increased expression or activity of HTRMs. HTRM
CC polypeptides, their immunogenic fragments or oligopeptides are useful for
CC screening libraries of compounds in drug screening techniques.
CC Polynucleotides encoding HTRM are useful for blocking the transcription
CC of mRNA and regulating gene function by modulating the activity of HTRM.
CC Vectors expressing HTRM or agonists can also be used to prevent or treat
CC disorder associated with decreased HTRM expression. Antibodies which
CC specifically bind HTRM and polynucleotides encoding HTRM are useful for
CC diagnosing disorders associated with the expression of HTRM, particularly
CC in assays that detect the expression of HTRM. Nucleotide sequences
CC encoding HTRM may be useful to generate hybridization probes useful in
CC mapping the naturally occurring genomic sequence and to detect
CC differences in gene sequences among normal, carrier and affected
CC individuals. Using diagnostic assays, cancer can be detected prior to their
CC appearance of clinical symptoms and thereby progression of cancer can be
CC prevented by aggressive treatment or preventive measures.

SQ Sequence 257 AA;

Query Match	Score	DB	Length
12.48	479	21	257

Matches 103; Conservative 50; Mismatches 73; Indels 20; Gaps 7;

QY 251 LTGAGVSYSVCGIPDPRS-RDGIYARLAVDPDLPDQAMFDEIFRRKDRPFKEFAKEIY 309
: |||: | ||||| :|: | |||: :| :|:| | |||
Db 1 mnyagaglstpsgipdrftrpsgslslnl--qqydlpypeatfelpffhnpypfftlakely 58

QY 310 PGGFQPSLCHKFIASLSDKEGKLLRNNTQNIIDITEQVAGI--QRIIQCHGSATASCLICK 367

Db 59 pgnqykpnrchylflllhdkglllrlyqniidgervsqpsaklveahngftasatclvcq 118

```

QY 368 YKVDCEAVRGDIFNVVPRRCRPADEPLAIKMPETVFFGENTPEQF--HRAMKYKDEV 425
      | : | : | | | | | : | : | : | | | | | : | : |
Db 119 rpfpgedlradmadrprpcvc-----tgvvkpdvlfgeplqrfllh---vvdipma 170

```

QY 426 DLIIVIGSSILKVRPALIPSSIPHEVQILINRPLPHLF-----DVLLIGDCVIINE 480
 |||:::|::| : : : ||::|::| : : | | ||
 Db 171 dIIIIIGTSLEVEPTASLTAEVRSSVPRLLINRDLVGPLAWHPRSRDAVQGDVNVGES 230

QY	481	LCHRLG	486
Db	231	lve11g	236

RESULT	2
Y48540	
ID	Y48540 standard; Protein; 212 AA

AC Y48540;

DT 08-DEC-1999 (first entry)

Human breast tumour-associated protein 1.

kw Expressed sequence tag; EST; human; breast; cancer; gene therapy;

XX
XX

OS Homo sapiens

XX	DE19813839-A1.
PN	

PD 23-SEP-1999.

PF 20-MAR-1998; 98DE-1013839.

PR 20-MAR-1998; 98DE-1013839.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;

DR WPI: 1999-528981/45.

XX

PT tissue, useful for breast cancer therapy -

PS Claim 25; 143; 188pp; German.

This invention describes novel human nucleic acid sequences from tumor breast tissue which have cytostatic activity. The nucleic acid sequences can be used to produce and isolate full-length gene sequences. They can be used to express proteins, which can be used as tools to find an activity against breast cancer. The sequences can be used in sense or antisense form. They are especially useful for medicaments for gene therapy to treat breast cancer. Y85350.1Y8617 represent protein fragments encoded by the expressed sequence tags described in the method of the invention.

Sequence	212 AA;
SD	

Query Match	9.28;	Score 355.5;	DB 20;	Length 212;
-------------	-------	--------------	--------	-------------

Matches 73; Conservative 35; Mismatches 49; Indels 7; Gaps 4;

Qy 206 ELDDMTLQIVINILSEPPKKRRKRDINTIEDAVKLLQ--ECKKIIVLTGAGVSVSCGIP 263

Dy 264 DRRS-RDGIYARLAVDEPDLDPQAMEDIETFFKDDPRPFKFAYELYPGQPSLCCHKFI 3222

Dd 102 dfrrspstclaydnl--ekyhnpypaeateisysfkfhpepfalakealypgqfptichyfml 1555

```

QY      323 ALSDKEGKILRNNTQNDITLEQVAGIQR--ILQCHGSFATASCL 364
      |  :| ||| |||||::|:::  :: ||| | : | :
Db      160 rllkdqgllrcytcnidtleriaqlgedlveahgtyftshcv 203

```

RESULT	3
W98718	
ID	W98718 standard; Protein; 205 AA

AC W98718;

DT 31-MAR-1999 (first entry)

DE H. pylori GHP0 808 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX

XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700

XX

XX

[illegible]


```

Db 954 esqdsq-----ksledes--getufgplekene-----slrslagqdqeeqklege 997
Qy 723 TRQELTDV 730
   1:1 1 1
Db 998 tqqlrlav 1005

RESULT 5
R60126
R60126 standard; Protein: 1805 AA..
XX
XX R60126;
XX
XX 21-MAR-1995 (first entry)
XX
XX Rat nestin protein is useful to identify brain tumours.
XX
XX nestin gene; brain tumour; neoplastic cells; glial; neuronal;
XX muscle; neural multipotential stem cell; mammalian brain; detection;
XX diagnosis; medulloblastoma; gliablastoma; oligodendroglioma.
XX
XX Rattus rattus.
XX
XX
XX Key Location/Qualifiers
XX MISC-difference 167..168
XX FT /note="encoded by CACG"
XX FT MISC-difference 172..173
XX FT /note="encoded by CCG G GCC"
XX
XX US5338639-A.
XX
XX 16-AUG-1994.
XX
XX 12-APR-1988; 88US-0180548.
XX
XX 12-APR-1988; 88US-0180548.
XX 02-JUN-1988; 88US-0201762.
XX 25-OCT-1990; 90US-0603803.
XX 22-FEB-1991; 91US-0660412.
XX 19-MAR-1992; 92US-0853913.
XX
XX (MAST ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Lendahl U, McKay RDG;
XX
XX MPI: 1994-263332/32.
XX N-PSDB: Q70447.
XX
XX Nucleotide and protein sequences for human and rat nestin -
XX distinguishes neural multipotential stem cells and brain tumour
XX cells from more differentiated cell types; for use in the
XX diagnosis of brain tumours
XX
XX Claim 4; Column 35-44; 45pp; English.
XX
XX This sequence is the rat nestin protein encoded by Q70447. Its
XX degree of sequence homology to intermediate filaments in the core
XX domain, its dissimilarity in the head and tail domains, and its
XX different splicing pattern suggest that it defines a new class of
XX intermediate filament protein. Nestin protein expression distinguishes
XX neural multipotential stem cells and brain tumour cells from the more
XX differentiated neural cell types (eg., neuronal, glial and muscle
XX cells of the adult brain). The nestin protein can be used in diagnosing
XX tumours of the brain, such as medulloblastomas, gliablastomas and
XX oligodendroglioma. (See also R60127).
XX
XX Sequence 1805 AA;
XX

```

QY	41	PELGNSPEEPSSAAVPAAGGEAASAAAPALMTPEAGAASA---	BREAPA-----	89
Db	332	PLISEAPCPPIEAATRAAGEVPSLLQTGPEPIWIKATVPSSSALLPEELGPGKqgghfP	451	
QY	90	-----TAVAGCDGNGSGARRPRRADDFDDECE-----	EEDEAATAA 127	
Db	452	ddltslatlnlphlphtleakgess-----	esrYsslfqdeegqlwelveadeievke 506	
QY	128	AAAAIGYADNLLIDGLLTNGPHSGCESDDDDRTSHASSDWTTPRPIGPTFYVOOHLM--	185	
Db	507	nsasqkfqesqldte-----	etqdsq-----gp---lqkelka 537	
QY	186	IGTDRTILKLDLPETIRPELDMWTLMQIYINLTSEPPKPK-----	KRKDI---NTI 235	
Db	358	lgeepImskIdqnyetacgkencmsnc-eghllgtlegpekqkIplksleeknvesekI	595	
QY	236	EDAVALLDCECKKIIVLTNAGVSVSGCIPDFFSRGIYARLAVDPDPDPDPAMFDIEYFR	295	
Db	596	engvrvlsel-----	lgrkedrtted-----qelmspgtcl----- 625	
QY	236	KDPRPFEPFAKEIYPGOFOPSLCHKFTALSDCKEKLRLNTQNTIDTLEQVAGIORILLCH	355	
Db	626	---krfsslgke-----	qgevrvpkegn-lewtafkeesqhlpg----- 666	
QY	356	GSFATASCLICKKYKVDCAVAVGDIFNQVVPVPCRCPADPELATIMKPELVFGGEMLPQFH	415	
Db	663	---fpgaqdmlerlvekcd-----	qsfrpspeeddeacr-----lqkenq 702	
QY	416	RAMKTKDKE---VDLLIYIGSLKVRPALIPSSIPIHEVPQIL--INREPLPHLPVEL	470	
Db	703	eplygeeeegqllerlleekesqelsr-----	peedeqaagslqkenqeplygeee--- 755	
QY	471	LGDCDIITNELCHRLGEGYAKLCCNPVKLSITEKPPRP-OKELVHLSELPT-----	522	
Db	756	---qdmlierllekesqelsk---speengrIqkplerenqkslryleengeltvpliesr	808	
QY	523	-----PLH-----	ISDDSSPERTVQDSSVATATVDQATNN 554	
Db	809	nqrIrlsrleveeeegrlvklpleksqdsIgsIaeenqyrlryleeddcInksllledtkh	866	
QY	555	NVNDLE-----	VSESSCEVEKEPOEQTNRNENINVENPDRKAVGSSSTADKNER 603	
Db	869	slgsIedrngdsIllpgsesetq-vslrppedeqdrrlynhllekesqel-----	sr 916	
QY	604	TSVAETVRKCPNRLAKEQISKR-LEGNOYLFVPPNRYIFHGAEVYSDESDDVLSSSCG	662	
Db	917	ssee-----	eqvmerslegn-----hesIsvokedqmwesqlek 953	
QY	663	SNSDSGTQOSSLERPLEDESEIEFYGLDEDDTEREPBCAGSGSFAGDGGQEVVNAIA	722	
Db	954	esqdsq-----	ksIedes-qetIlgplekena-----slrsIagdqdeeqkIe 997	
QY	723	TROELTDV 730		
Db	998	tqgtlrav 1005		
RESULT	6			
ID	Y56967	Y56967 standard; protein; 1192 AA.		
XX	Y56967;			
XX	25-APR-2000	(first entry)		
DE	Human MAGI polypeptide.			
KW	MAGI protein; neuroendocrine-specific protein; neuropathy; human;			
KW	spinal injury; neuronal degeneration; neuromuscular disorder; cancer;			
KW	psychiatric disorder; developmental disorder; inflammatory disorder;			
KW	stroke; cytostatic; cerebroprotective; neuroprotective.			
OS	Homo sapiens.			

XX WO200005364-A1.
 PN
 XX 03-FEB-2000.
 PD
 XX 21-JUL-1999; 99WO-GB02360.
 PF
 XX 22-JUL-1998; 98GB-0016024.
 PR 19-JUL-1999; 99GB-0016898.
 XX
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 XX Michalovich D, Prinjha RK;
 PI
 XX WPI: 2000-182693/16.
 DR N-PSDB; 256886.
 XX
 XX Novel polypeptides related to neuroendocrine-specific proteins and
 PT polynucleotides useful for diagnosis of various diseases and for
 PT treatment of cancer and neurological disorders -
 XX
 XX Claim 2; Page 20-21; 35pp; English.
 PS
 XX The invention relates to human MAGI protein, which is similar to
 CC neuroendocrine-specific protein. The MAGI protein can be expressed by
 CC standard recombinant methodology. The MAGI polypeptides, polynucleotides
 CC and antibodies are useful for treating diseases, including neuropathies,
 CC spinal injury, neuronal degeneration, neuromuscular disorders,
 CC psychiatric disorders and developmental disorders, cancer, stroke and
 CC inflammatory disorders. The polynucleotide is also useful for chromosome
 CC localization and for tissue expression studies. The present sequence
 CC represents the human MAGI protein.
 XX
 SQ Sequence 1192 AA;

Query Match 3.4%; Score 132; DB 21; Length 1192;
 Best Local Similarity 18.9%; Pred. No. 0.015;
 Matches 176; Conservative 110; Mismatches 356; Indels 290; Gaps 38;

QY 12 ACSPSAAAMEASOPADPELRKRPRRDGGLGRSPESAAYAPAA-----ACCEAA 64
 DB 119 apspisaavspsklpdeddepparp-----ppppasvsgaepwtpapapaa 168
 QY 65 SAAAPALAREAAAGAA-----ASAE-----REAPXAV-AGCGDNGSG 102
 DB 169 pspstpaapkrirgsgsvdeltalpaasepvrissaeemdlkegpnrltsagdedfsvl 228
 QY 103 RR-----EPRAADDFDDDEGEDEEAAAAAAGVRDNLTLTGILTGFSCHS 154
 DB 229 letaaslpalslaasfkehe-----ylnlstvlptegltlgenveaak 274
 QY 155 D-----DDRTSHAS-----SSDWTPPRIRIGPYTFVQOHLMTGTPR--TILKD 196
 DB 275 evsekaktlldrdltfeseleysemgsfsfvsrk-----aesavlvampreelivn 327
 QY 197 -----LLEPFI-----PPPELDKMTLMQIYIN-----ILSEPPRRKKRKDI 232
 DB 328 kkeekelvannllhngqelftalclkvkedevvsekaakdfnekrvavaeapmreeyad 387
 QY 233 NTIEDAVKILQECKRIIVLTGAGVSVSGCIPD-----FRSRDGIYARLAVD 278
 DB 388 kfe-rvwevkskedsdmaagklesnleskvdckfadsleqtnheksessnddts 446
 QY 279 FPDLPD-----PQAMFIDFYFRKDRPPRFKFAKEIYPGOFPSLCHK----- 320
 DB 447 fstppegklrpgayltcapf--npatesiatnlfplgdpksenktdekkleekaq 504
 QY 321 -----FTALSDKECKLKNYTOND--TLEQVAG-----IQRLQCH 355
 DB 505 vteknstcktsnplfvaadsetdyvt--tdnltkveevvampegllpdlvqaecese 562
 QY 356 GSFATASCLICKYKVDCEAVRGDIFNQVVRPCRPCPAD-----PLAIKMPETIV--FG 407

DB 563 lnevtgklayekmdlvqtsewmgesltpaaqlcpsfeesaeatpvpvl-pdlvmeapln 621
 QY 408 ENLPPEQFHRAMKYDKDEVLLIYIGSSLKVRYPALP-----SSTPHEVPQL 455
 DB 622 saypsagaavilqpspspleassvnyeslkhepenpppyeamsvsllkvsglkeel---- 677
 QY 456 INNEPLPHLFVDVELLDGDDVITINELCHRLGGEYAKLCNPNVLSBIT----- 503
 DB 678 --kep-----eninaalqetapylsiacdlkckliasepapdfdsyse 720
 QY 504 ----EKPPRQKELVHLSELPTPLHISEDSSPERTVPDSSVATLVLDQATNNVNDL 559
 DB 721 makveqpydhselvedspsdsepvdlfddslpd--vpqkqdetvmlkseit----- 772
 QY 560 EVSESSCVEEKPOEVOTSRNVENINVENPDPKAVSGSTADKNERISVAETKRCWPNRLA 619
 DB 773 etsfeemleyenkeklsalppegkpylesfkislndtcltlpdevs--tlskkekplq 831
 QY 620 KEOISKRLGNOYLFVPPNRYIFHGAEVYSDS-----EDVLSSSSGGS----- 663
 DB 832 meelstaavsndllfskdaqf-retetltdssapleidefptllsktdsfklareyt 890
 QY 664 -----NSDSCGTCSPSLEBPLE-----DESEIEEFYNGLEBDTEREPCAGSGF 707
 DB 891 dlevshkselanapdgaglpctelphdlslnkdpkveeksfsdd-----f 938
 QY 708 GADGGOQEVV-----NEAIAIRQELTDVNY 733
 DB 939 skngsatskvlllpdvlsalatqaeleslvkp 970

RESULT 7
 ID W56163 standard; Protein; 738 AA.
 XX
 AC W56163;
 XX
 DT 28-JUL-1998 (first entry)
 XX
 DE New DNA sequence isolated from Pinctada fucata.
 XX
 KW Pinctada fucata; protein; cosmetic.
 XX
 OS Pinctada fucata.
 XX
 PN JP10080285-A.
 XX
 PD 31-MAR-1998.
 XX
 PF 28-MAY-1997; 97JP-0138461.
 XX
 PR 15-JUL-1996; 96JP-0184459.
 XX
 PA (MIKI-) MIKIMOTO SEIYAKU KK.
 XX
 DR WPI: 1998-254410/23.
 XX
 DR N-PSDB; V22683.
 XX
 PT New cDNA and e.g. vector, host cell and polypeptide - used to
 PT produce polypeptide in high yields, which is used in cosmetics
 PS Claim 9; Pages 9-11; 15pp; Japanese.
 XX
 CC The present sequence represents protein encoded by a new DNA sequence
 CC isolated from Pinctada fucata. The protein be used as an ingredient
 CC in cosmetics.
 XX
 SQ Sequence 738 AA;

Query Match 3.2%; Score 122; DB 19; Length 738;
 Best Local Similarity 32.3%; Pred. No. 0.053;

Matches 61; Conservative 9; Mismatches 65; Indels 54; Gaps 8;

QY 5 VALALQAG-----SPSAAMEASQPADEPLRKRRPDRDGPGLGR 45
 Db 481 laaalaagaagggfaglgaglgaggsaaaaaaaaaagggatrratrrtmrggs 540
 QY 46 SPGEPSAANA-----AAGEAASAAAPALMREAGAAA 81
 Db 541 aaaaaaaaaagggwggmggfgvlgggfaggfagggsaaaaaaaaaagfgg 596
 QY 82 SAEEAATAVAGDGGSGLRREPRADPDDEGEEDPAAAAAAGYDNLILT 141
 Db 597 gg-rrrgrrgrrgdd-gngasavaaaaaaaaaaaggaadvaaaaaaam-ygdy--a 650
 QY 142 DGL-LTNGF 149
 Db 651 dgpfdingf 659

RESULT 8
 ID Y53970 standard; Protein; 1857 AA.
 XX Y53970;
 DT 13-MAR-2000 (first entry)
 XX
 DE Human peripheral benzodiazepine receptor associated protein-1.
 KW Human; peripheral benzodiazepine receptor associated protein-1;
 KW PRA-X-1; peripheral benzodiazepine receptor; chromosome 17;
 KW central nervous system; immune system; gene therapy;
 KW PRA-X-1 deficiency condition; endocrine system.
 OS Homo sapiens.
 PN WO960117-A2.
 XX
 PD 25-NOV-1999.
 XX
 PF 06-MAY-1999; 99MO-FR01070.
 XX
 PR 15-MAY-1998; 98FR-0006190.
 XX
 PA (SNFI) SANOFI-SYNTHELABO.
 PI Casellas P, Gallegue S, Jbilo O, Le Fur G;
 DR WPI; 2000-062455/05.
 DR N-PSDB; 236989.
 XX
 PT New PRA-X-1 polypeptide that interact with peripheral benzodiazepine
 PT receptor, used to treat e.g. immune, central nervous or endocrine
 PT disorders -
 PS
 PS Claim 1; Page 39-44; 44pp; French.
 XX
 CC The present sequence represents a peripheral benzodiazepine receptor
 CC associated protein-1, designated PRA-X-1. PRA-X-1 interacts specifically
 CC with the peripheral benzodiazepine receptor. The PRA-X-1 gene is
 CC localised on chromosome 17 in the q22-q23 region. The gene is
 CC associated with markers of pathologies of the central nervous system
 CC or immune system. The PRA-X-1 nucleic acid is useful in gene therapy (of
 CC CC immune or endocrine systems; e.g. disorders of the central nervous,
 CC immune or endocrine systems; as a source of diagnostic primers and
 CC probes (see Z36990-Z37023) and of antisense therapeutics; for
 CC recombinant production of the PRA-X-1 protein; and for detecting allelic
 CC variants, mutations, deletions, insertions, loss of heterozygosity and
 CC gene rearrangements in the PRA-X-1 gene. The PRA-X-1 protein is used to
 CC raise specific antibodies and to screen for specific modulators
 CC (potential therapeutic agents). The antibodies are used as immunoassay
 CC reagents, e.g. for diagnosis of abnormal expression or accumulation of
 CC PRA-X-1.

XX XX
 SQ Sequence 1857 AA:

Query Match 3.28; Score 121.5; DB 21; Length 1857;
 Best Local Similarity 19.38; Pred. No. 0.25;
 Matches 171; Conservative 82; Mismatches 253; Indels 381; Gaps 46;

QY 2 ADEVALALQAGSPSAAMEASQPADEPLRKRRP-----RNDGPG----- 42
 Db 1067 adsiipalipalapasipatvscpspnspearplasaaspgpdpssplqhpapigtge 1126
 QY 43 -LGRSPGEPSAANA-----PAAAGCEAASA-----AAPAA 71
 Db 1127 ppgapppaspremakgshedppapcsgeaavlygtseertaststlgckdpgpaapsl 1186
 QY 72 LMRPA-----AGAAASAE-----REAPATAVAGDGNCSGLRREPPA----- 108
 Db 1187 akgeawtagaaccpassttgatagqapntemcggqgpgqglr--praekedaelyvhl 1244
 QY 109 -----ADDFDDDEGEEDPAAAAAAGYDNLILTDLTNGFHS----- 151
 Db 1245 vnsivdhgrnsldsidgeeeeeeeeeelgrtscfkyq-vagnsirengakspdpf 1303
 QY 152 CESDDDDRT-----SHASSD-WPRPR 173
 Db 1304 ceidsdeellleqllelplqgfcskllfsipeeeeeeeeeeksgagcsrdpgppepa 1363
 QY 174 IGRPTVQOHLMTGD---FRTILKDLLETIP-PPE-----LDMTLMQVYINL-- 220
 Db 1364 l-----lgldcsdqgqrr-----pqcpbsspsrsgdledmp-----glvgs 1402
 QY 221 -----SEPRKRRKDINTIEDAVKILQCKKITIYVTGAGVSVSGIPDFRSDG 270
 Db 1403 ssrrggsgspekpsrrppdr-----elcsl-lsmngpaasgrlpgtrerg 1451
 QY 271 IYARLAVDPDLDPQAMFIE-YFRKDPREFFAKEIYPGOFPSLCHKFNALS--D 326
 Db 1452 l-----pvlegrrtgaasgrlgrprcrrcgralerg-lasclspkcleisleyd 1502
 QY 327 KEGKLLKNYTONIDTLEQVAGIQRILOCHGSFATASCLICKYKVCDA-----VRGD 378
 Db 1503 sed-----eqeag-----sgglstlsc-----ypgdreaawgatvgrprp 1539
 QY 379 IFNQVPR-CPRCPADPEPLAIIMKPEIVFEGENLPEQHR-AMKVDKDEVOLLIVIGSS-- 434
 Db 1540 pkansgpkyprrlpawek-----ge--perrgrsatgrakeplisratetgearg 1586
 QY 435 -----LKVREVALIPSSIPHEVPQILINREPLPHLFDVEL---LQDCD 475
 Db 1587 qdgsgrtgpqkryrvlrrpstaelyvparpsset-----layqlpvrilfvalfdyd 1637
 QY 476 VIT-----NRLCRL-----GGEYAKKLCNRYKALSEIRKEP 506
 Db 1638 pvsmspnpdageeelpifregljkvfgkdadgfygggggrrtgyipcmnv--aeavnds 1695
 QY 507 PRPOKELY-----HLSEPPPLMLHSE-DSSSPERTVPDS 541
 Db 1696 pagtqgllqgyispdlillegsgnqpfvystahltgppprrskkeesegpqpccgpp 1755
 QY 542 SVATLVDQATNNNVNDELEWSES--SCVEEKPOEVOGTSRNV----- 581
 Db 1756 klvpasa-----dlkaphsmvaafdynpgsspmndvaelprtagdvltvfgtd 1805
 QY 582 -----NIN-----VENPDFRAVGSSTADKKERTSVAEVTR 611
 Db 1806 ddfgygelngqrglypsnflfegpgeag--ldreptlrgaesqr 1849

RESULT 9
 Y40098
 ID Y40098 standard; Protein; 531 AA.
 XX


```

AC  Y40098;
XX
XX  19-NOV-1999 (first entry)
XX
XX  Spider silk protein spidroine major 2.
DE
XX
XX  Spider silk protein: spidroine major 2; cosmetic; make-up;
KM  dermatological compositions; hair care; skin care; sunscreen;
XX  hormone; moisturizer; skin disorder; skin disorder.
OS  Nephtila clavipes.
XX
XX  FR2774588-A1.
XX
XX  13-AUG-1999.
XX
XX  11-FEB-1998; 98FR-0001614.
XX
XX  11-FEB-1998; 98FR-0001614.
XX
XX  (OREA ) L'OREAL SA.
XX
XX  Philippe M, Garson JC, Arraudeau JP;
XX
XX  WPI; 1999-510729/43.
XX
XX  Cosmetic or dermatological composition containing spider silk protein,
XX  for hair or skin care, in make-up or sunscreens
XX
XX  Claim 5; Fig 2; 32pp; French.
XX
XX  The present sequence represents the natural spider silk protein
XX  spidroine major 2. The protein improves the moisturizing/softening
XX  action of the compositions. The protein, and its fragments are used
XX  in cosmetic or dermatological compositions. These compositions have
XX  use as hair or skin care products; and make-up or sunscreens.
XX  As the protein is a good, persistent film-formers on the skin
XX  of low surface density, it can be used for delivery of active
XX  agents that are generally difficult to administer, e.g. vitamins,
XX  hormones, moisturizers or agents for treating disorders of the
XX  skin and hair.
XX
XX  Sequence 531 AA;
SQ
Query Match 3.1%; Score 121; DB 20; Length 531;
Best Local Similarity 34.9%; Pred. No. 0.039;
Matches 45; Conservative 8; Mismatches 64; Indels 12; Gaps 4;
QY 12 AGSPSAAAMFASQ-----PADE-PLKRPRRDGPGICRSPGPSAAVAPAAACEAA 64
DB 308 agsaaaaaaagpgqg1g9y9pgq9pg9y9pgq9pg9-9y9pgsaaaaaaagpgq9gp 366
QY 65 SAAPALMREAGAAASAREAPATVAGDNGSGILRREPRRAADFDDEGEEDDEAA 124
DB 367 g9y9pgq9pgspsasasa-----aaaaaaagpg9y9pgq9pg9y9pgq9pgspsasaa 422
QY 125 AAAAAAAG 133
DB 423 aaaaaagpg 431
DE
XX  RESULT 10
XX  R14309 standard; Protein; 595 AA.
XX
XX  R14309;
XX
XX  15-JAN-1992 (first entry)
XX
XX  N.clavipes dragline silk protein-2.
XX
XX  protein superfibre; major ampullate silk; orb web spider.

```

```

XX
XX  Nephtila clavipes.
XX
XX  EP452925-A;
XX
XX  23-OCT-1991.
XX
XX  18-APR-1991; 91EP-0106217.
XX
XX  20-APR-1990; 90US-0511792.
XX
XX  (UYWY-) UNIV OF WYOMING.
XX
XX  Lewis RV, Xu M, Hinman M;
XX
XX  WPI; 1991-312199/43.
XX
XX  N-PSDB; Q14184.
XX
XX  DNA encoding spider silk protein-1 and 2 and variants - isolated
XX  from Nephtila clavipes, for prodn. of spider silk protein and
XX  fibres having desired characteristics
XX
XX  Claim 16; Page 30; 48pp; English.
XX
XX  The spider silk protein contains repeat regions comprising an
XX  amorphous region which can form an alpha-helix when stretched and a
XX  region which can form beta-pleated sheets. Removal of the poly-(Ala)
XX  segments results in a silk having lower elasticity. It is envisaged
XX  that derivatives of the silk protein could be produced having
XX  different tensile strength or elasticity than the natural protein
XX  by altering the ratio of different amino acids.
XX  See also Q14183 and Q14185.
XX
XX  Sequence 595 AA;
SQ
Query Match 3.1%; Score 121; DB 12; Length 595;
Best Local Similarity 34.9%; Pred. No. 0.046;
Matches 45; Conservative 8; Mismatches 64; Indels 12; Gaps 4;
QY 12 AGSPSAAAMFASQ-----PADE-PLKRPRRDGPGICRSPGPSAAVAPAAACEAA 64
DB 308 agsaaaaaaagpgqg1g9y9pgq9pg9y9pgq9pg9-9y9pgsaaaaaaagpgq9gp 366
QY 65 SAAPALMREAGAAASAREAPATVAGDNGSGILRREPRRAADFDDEGEEDDEAA 124
DB 367 g9y9pgq9pgspsasasa-----aaaaaaagpg9y9pgq9pg9y9pgq9pgspsasaa 422
QY 125 AAAAAAAG 133
DB 423 aaaaaagpg 431
DE
XX  RESULT 11
XX  W53347
XX  ID W53347 standard; Protein; 595 AA.
XX
XX  W53347;
XX
XX  06-JUL-1998 (first entry)
XX
XX  Nephtila clavipes spider silk protein.
XX
XX  Spider; Nephtila clavipes; silk protein; tandem repeat; fibre; dragline;
XX  cocoon; tensile strength; elasticity.
XX
XX  Nephtila clavipes.
XX
XX  US5728810-A.
XX
XX  17-MAR-1998.
XX
XX  19-APR-1995; 95US-0425069.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:45:30 ; Search time 80.58 Seconds

(without alignments)
621.032 Million cell updates/sec

Title: US-09-461-580A-26

Perfect score: 3854

Sequence: 1 MADEVALALQAGSPSAAAA.....NEAATRQELTDVNPSPDKS 737

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_66:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	923.5	24.0	607	2	T24172	hypothetical prote
2	621.1	16.1	670	2	S36616	regulatory protein
3	605.5	15.7	471	2	T39571	probable regulator
4	592	15.4	503	2	S59698	HST1 protein - yea
5	568	14.7	562	1	RGYS2	regulatory protein
6	538	14.0	331	2	T18247	transcription regu
7	513.5	13.3	332	2	T40929	transcription regu
8	484	12.6	381	2	UC4639	silent information
9	439	11.4	357	2	S59678	HST2 protein - yea
10	349	9.1	247	2	H69827	conserved hypotet
11	346	9.0	246	2	A72370	regulatory protein
12	341.5	8.9	447	2	S54631	HST3 protein - yea
13	330.5	8.6	253	2	H69263	transcription regu
14	295	7.7	245	2	C69459	transcription regu
15	285.5	7.4	247	2	D72562	hypothetical prote
16	282.5	7.3	249	2	F71085	hypothetical prote
17	267.5	6.9	250	2	C75101	transcription regu
18	267	6.9	250	2	A83506	probable cobalamin
19	256	6.6	415	2	T50106	hslap [imported] -
20	245	6.4	239	2	C70486	conserved hypotet
21	243.5	6.3	370	2	S52699	hypothetical prote
22	220	5.7	287	2	T23234	hypothetical prote
23	211	5.5	237	2	H70554	hypothetical prote
24	210	5.4	259	2	H82191	probable nicotinat
25	210	5.4	279	2	E64856	hypothetical prote
26	200.5	5.2	256	2	A83495	hypothetical prote
27	197.5	5.1	234	2	A71838	hypothetical prote
28	190	4.9	287	2	T23325	hypothetical prote
29	187.5	4.9	233	2	G81307	probable transfera

30	186	4.8	3534	2	T42567	tegment protein 2
31	185.5	4.8	246	2	G75570	conserved hypotet
32	179	4.6	1095	2	T25520	hypothetical prote
33	158.5	4.1	190	2	T33951	probable STR2 faml
34	150.5	3.9	3421	1	W2BE86	367K tegument prot
35	147.5	3.8	460	2	T33110	hypothetical prote
36	146	3.8	118	2	T46348	hypothetical prote
37	146	3.8	2484	2	T26216	hypothetical prote
38	146	3.8	2607	2	T26215	hypothetical prote
39	143.5	3.7	2639	2	T31328	hypothetical prote
40	142	3.7	1137	2	G70868	fibroin - Chinese
41	138	3.6	205	2	C64677	probable regulator
42	135	3.5	1805	2	A34736	conserved hypotet
43	134.5	3.5	587	2	B70884	nestin - rat
44	134	3.5	1857	2	T50513	hypothetical prote
45	133	3.5	1906	1	S68235	myosin-11gnt-chain

ALIGNMENTS

RESULT	1	24.0%	Score 923.5	DB 2	Length 607
T24172	hypothetical protein R11A8.4 - Caenorhabditis elegans	Best Local Similarity 36.5%	Pred. No. 1.5e-50		
C:Species: Caenorhabditis elegans		Matches 217	Conservative 106	Mismatches 186	Indels 85
C:Date: 15-Oct-1999	#sequence_revision 15-Oct-1999				Gaps 15
C:Accession: T24172					
R: Cummings, P.					
A: submitted to the EMBL Data Library, March 1996					
A: Reference number: Z19849					
A: Accession: T24172					
A: status: preliminary; translated from GB/EMBL/DBJ					
A: Molecule type: DNA					
A: Residues: 1-607 <MIL>					
A: Cross-references: EMBL:270310; PIDN:CAA94364.1; GSPDB:GN00022; CESP:R11A8.4					
C: Experimental source: clone R11A8					
C: Genetics:					
A: Gene: CESP:R11A8.4					
A: Map position: 4					
A: Introns: 18/3; 68/3; 111/3; 206/3; 500/3; 565/3					
Query Match					
Best Local Similarity	24.0%	Score 923.5	DB 2	Length 607	
Matches 217	Conservative 106	Mismatches 186	Indels 85	Gaps 15	
QY 158 DRTSHASSDWTFRPR-IGPYTVQOHMIGTPRTILKDLPE--TIPPELDDMTLMQ 214					
DB 55 ESTTSSSESMQNNDEMNSLRRAQRLLDGGATPRLQIQIPDFDNMSRIATMSEMHFA 114					
QY 215 IVNIISEPPKRRKRKDINTIEDAVKILQECCKIIVLTGAGSVSCGIPDFRSRDGIYAR 274					
DB 115 ILLDLERAVEROKLTNYNSLADAVELFKRKHILVLTGAGSVSCGIPDFRSKDGIIYAR 174					
QY 275 LADPEPLDPQAMFDEYERKDPREFRAKTEIYPOQPSLCHKRIALSDMEGKILRN 334					
DB 175 LSEEPDLPPPTAMFDIYRFPNPAPYFNPARETFPQFVPSVSHRIKELFSGRLRN 234					
QY 335 YTONIDLEQVAGIQRILQCHGSFATASCLICKYKVCSEVRSDIFNOVPRPCPAD 394					
DB 235 YTONIDLEQVAGIQRILQCHGSFATASCLICKYKVCSEVRSDIFNOVPRPCPAD 394					
QY 292 --GVIRKNIVFEGEDLGRFQHVTEDKHVYDLIVIGSSLKVRVALIPIHCDKNVPOI 349					
QY 395 PLAIIMKEIIVFEGENLPEQFHRAKKYDKDEVDLLIYIGSSLKVRVALIPIHCDKNVPOI 454					
DB 455 LIRREPLPHLAPVELLDGDDVYIINELCHRLGGEYAKLCNPKLSP-----ITERPPR 509					
QY 350 LIRREPLPHLAPVELLDGDDVYIINELCHRLGGEYAKLCNPKLSP-----ITERPPR 509					
QY 510 OKELVHLSLEPPLPLH-----SEDSSEPPERTYPODSVYATLVADQATNNVNDLEVES 564					
DB 410 KRLLISGEDF-----LNICMKERKNDSSDEPTLKKPRMSVA-----DSMDSEKNNF----- 457					

```

OY 565 SCVEERKPOVOVSRAVNEVNVENPEFKAVGSSADTKNEFTSVAEVVRKQWPRRLKEDIS 624
      ||| : : : : : |||
Db 458 -----OEIQHKSEDDDDTRNSD-----DILKIKIPRL--SIT 490
      ||| : : : : : |||
OY 625 KRLGNOYLVPBNRYIFGAE-----VYSDSEDDVLSSSCGSSNSDGTQ 671
      ||| : : : : : |||
Db 491 EMLHNKCAVAISAHQTVFPGACSFDELTKLVRDVIHHETHCSCSGSSNADSEANQ 550
      ||| : : : : : |||
OY 672 ---SSLEP-PLDESEIEFYNGLEDTEREPCAGGSGFGADGSDQEVNNAI 721
      ||| : : : : : |||
Db 551 ISRAOSLDEVDLDEDDR---KNTHLDLQRAD-----SCDGFQYELSETI 593
      ||| : : : : : |||

RESULT 2
536616
regulatory protein SIR2 - yeast (Kluyveromyces marxianus var. lactis)
N:Alternate names: silent information regulator 2
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
C:Accession: A56048; S36616
R:Chen, X.J.; Clark-Walker, G.D.
Mol. Cell. Biol. 14, 4501-4508, 1994
A:Title: sir2 mutants of Kluyveromyces lactis are hypersensitive to DNA-targeting drugs
A:Reference number: A56048; MUID:94277055
A:Accession: A56048
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-670 <CH2>
C:Superfamily: regulatory protein SIR2
C:Cross-references: EMBL.X74569; NID:g397349; PIDN:CA52661.1; PID:g397350
C:Keywords: DNA binding; nucleus; transcription regulation

```

Query Match	Similarity	Score	DB 2:	Length
Best Local	29.58%	Pred. No. 1.8e-31;		
Matches 199:	Conservative	95;	Mismatches 196;	Indels 184;
				Gaps 26;

```

QY      436 KVRPALIPSSIPHVPOLLINREPLPHLHFPVELLGCDVIYNELCHRIAGE-----YA    490
           ||||| : |||||:::||| : || : || : |
Db       531 KVAPVSEIVNMLPAHVPOVLINKDPVKHAEPDLSGLGDVAALVAOKCGMDIPTHDMNN    590
                                           ::||| : || : || : |
QY      491 KLCCPVP-----KLSEITEKPPRPQ-KELYHL-----SELPTPLHISED---   529
           ||||| : ||| : || : || : |
Db       591 KL-KKKVFDSEVERGVYKLVHPINESPALEAEKEKHLPLOOSTAALTLP-PVSLSDSPG    648
                                           ::||| : || : || : |
QY      530 ---SSPERTVPQD 540
           ||||| : ||
Db       649 RRSSSSPQPRTQT D 662
                                           ::||| : ||

RESULT      3
            T39571
             probable regulatory protein sir2-like - fission yeast (Schizosaccharomyces pombe)
             C.Species: Schizosaccharomycos pombe
             C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
             C.Accession: T39571
             R.Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.
             submitted to the EMBL Data Library, March 1999
             A.Reference number: Z21864
             A.Accession: T39571
             A.Status: preliminary; translated from GB/EMBL/DDBU
             A.Molecule type: DNA
             A.Restrictions: 1-471 <WOO>
             A.Cross-references: EMBL.AL035657; PIDN:CAB38511.2; GSFDB:GN00067; SPDB:SPEC16D10.07c
             A.Experimental source: strain 972h-; cosmid c16d10
             C.Genetics:
             A.Gene: SPDB:SPBC16D10.07c
             A.Map position: 2
             A.Introns: 53/L; 79/L; 106/3; 117/2; 263/L

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:52:36 ; Search time 42.08 Seconds

(without alignments)
565.607 Million cell updates/sec

Title: US-09-461-580A-26

Perfect score: 3854
Sequence: 1 MADEVALALQAGSPSAAAA.....NEAIATROELDVNYPDSKS 737

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	621	16.1	670	1	SIR2_KLULA
2	592	15.4	503	1	HST1_YEAST
3	568.5	14.8	515	1	SIR2_CANAL
4	568	14.7	562	1	SIR2_YEAST
5	484	12.6	381	1	SIR2_LETMA
6	439	11.4	357	1	HST2_YEAST
7	341.5	8.9	447	1	HST3_YEAST
8	243.5	6.3	370	1	HST4_YEAST
9	150.5	3.9	3421	1	TEG4_HSVB
10	135	3.5	1805	1	NEST_RAT
11	133	3.5	1906	1	KMLS_CHICK
12	130	3.4	3828	1	TRX_DROVI
13	128.5	3.3	1603	1	PSC_DROME
14	124.5	3.2	3969	1	HRX_HUMAN
15	124	3.2	1487	1	ICP4_HSVB
16	124	3.2	1487	1	ICP4_HSVB
17	121	3.1	627	1	SPD2_NEFCL
18	119	3.1	1184	1	FBL2_HUMAN
19	118	3.1	450	1	CYL_PARDE
20	118	3.1	627	1	YK09_CAEEL
21	117.5	3.0	739	1	BAC1_MOUSE
22	116.5	3.0	482	1	YSR2_CAEEL
23	116.5	3.0	606	1	HMD1_DROAN
24	116	3.0	814	1	IF39_HUMAN
25	113.5	2.9	386	1	HXAD_HUMAN
26	113.5	2.9	388	1	HXAD_HUMAN
27	113	2.9	705	1	ICAL_BOVIN
28	112.5	2.9	671	1	CHS5_YEAST
29	112.5	2.9	1676	1	APSA_EMENT
30	112	2.9	817	1	YG4A_YEAST
31	111.5	2.9	331	1	MACS_BOVIN
32	111.5	2.9	676	1	ICP0_HSVBK
33	111.5	2.9	1616	1	P200_MYCGE

34	111	2.9	1744	1	TANA_XENLA	001550 xenopus lae
35	110.5	2.9	908	1	SRCA_RABIT	P13666 oryctolagus
36	110.5	2.9	1505	1	CDP_HUMAN	P39880 homo sapien
37	110.5	2.9	3396	1	PGCV_HUMAN	P13611 homo sapien
38	109.5	2.8	563	1	ARX_MOUSE	O35085 mus musculu
39	109	2.8	88	1	H82_NEIGO	P11910 neisseria 9
40	109	2.8	226	1	BASP_HUMAN	P80723 homo sapien
41	109	2.8	465	1	FXD1_HUMAN	O16676 homo sapien
42	109	2.8	978	1	SIR3_YEAST	P06701 saccharomyc
43	109	2.8	1850	1	VIT2_CHICK	P02845 gallus gall
44	108.5	2.8	331	1	MACS_HUMAN	P29966 homo sapien
45	108.5	2.8	375	1	SOX3_MOUSE	P53784 mus musculu

ALIGNMENTS

RESULT	ID	STANDARD	PRT	670 AA.
1	SIR2_KLULA			
AC	P33294			
DT	01-FEB-1994 (Rel. 28, last sequence update)			
DT	01-FEB-1994 (Rel. 28, last sequence update)			
DT	01-NOV-1997 (Rel. 35, last annotation update)			
DE	REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).			
GN	SIR2.			
OS	Kluyveromyces fragilis (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;			
OC	Saccharomycetaceae; Kluyveromyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CHS 2359/152;			
RX	MEDLINE=94277055; PubMed=8007956;			
RA	Chen X.-D., Clark-Walker D.G.;			
RT	"SIR2 mutants of Kluyveromyces fragilis are hypersensitive to DNA-			
RT	targeting drugs.";			
RL	Mol. Cell. Biol. 14:4501-4508 (1994).			
CC	- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM			
CC	MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS			
CC	INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI			
CC	SILENCING (BY SIMILARITY).			
CC	- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	- SIMILARITY: BELONGS TO THE SIR2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X74569; CAA52661.1; -			
DR	PIR; S36616; S36616.			
KW	Transcription regulation; Repressor; DNA-binding; Zinc-finger;			
KW	Nuclear protein.			
FT	DOMAIN 44 94			
FT	DOMAIN 310 554			
FT	2N-FLING 428 455			
FT	DOMAIN 650 654			
FT	SEQUENCE 670 AA: 74227 MW: 00A063C/43308B75 CRC64;			
QY	Query Match 16.1%; Score 621; DB 1; Length 670;			
QY	Best local similarity 29.5%; Pred. No. 2.1e-30;			
QY	Matches 199; Conservative 95; Mismatches 196; Indels 184; Gaps 26;			
QY	32 LRRKPRDGLGSLGSPSPSAVAAPAAACEAASAA-----APALMLREAGAAASAE 84			
QY	8 LKRPRL-----QESVAGNGGLESGLKARGDSVFAARSPENEDVDADAD 55			
QY	85 REAATAVAGDNGSLRREPRAADFD--DECEDED----- 121			

```

Db 56 VDADADADADEDAOKDILEETKADELDEVDEEKEVSSNFGTASDHVGTSSMTG 115
Qy 122 -EAAAAA---AAGYRDNLTLTGTLNGFHCESDD---DRTS-----161
Db 116 STALLASSADNNGSGNGTGMATNGTSLDRYAPQKPEHPKLERRSYKRYEVPVSK 175
Qy 162 ---HASS-----SDWTPRRIGPYTVOOH--MIG---TDPTIL-----194
Db 176 EDSLARSYLQFGSARELDLPEDLNSLYY---HMIKLLGFGIKKELMALAQEYVH 232
Qy 195 ---KDLPE-----TIPPELDMTLMQIYINILSEPPK-----RK 227
Db 233 NADNDSPQKNSSETKVNSDTYATYPSFSEPLEKKHAARLTKDLQKANKVLSRI 292
Qy 228 KRKDINTEDAVKLLQECKKIIVLTGAGVSCGIPDRSRDGIYARLAVDPDLDPQA 287
Db 293 RLTNHTIDDPFAKLKTKAKIIVLGTAGISTSLGIDFRSSSGEYFKSG-DL-GIANDQD 350
Qy 288 MEDIEYFRKDRPPEFKAKIYVGOFPDPLCHKFTALSDKEGKLLRNTQNTIDLEQVAG 347
Db 351 VFSLVEFTEDSVFYNIAMVLPENMYSPLHSFIKMIQDKKLLRNTQNTIDNLESYAG 410
Qy 348 I--QRILOCHSEFATASCLICKYKNDCEAVKNDINOVVPRCPRC-----PADEP 395
Db 411 VEPEKMWCHSGFATASCVTCHMKITGERIFPNINQLPLCPYCYSKRLEFEKTKTDEE 470
Qy 396 LA-----IMKEIVFEGENLPEQFHRAMKYDKDEVDLLIVIGSSL 435
Db 471 LADGEDDMDDHGRSVKSKVLPDLITFGEALPSPKFRHRIREDVLAQCDLLICIGTSL 530
Qy 436 KVRPALIPSSIPHEVPOILNREPLPHLDVLLGDCVYIINELCHRLGE-----YA 490
Db 531 KVAPSEIVNMIPAHVPOLINKDPVKAHEFDLSLGLCDVAALVAQKCGWDIDPHDMN 590
Qy 491 KLCQPV-----KLSEITEKPPRPQ-KELVHL-----SELPPTLHISED---529
Db 591 KL-KKRVKVDSEVERGVYVHPLNESPALLEEKHLPLQOSTALTP-PVSLSADSPG 648
Qy 530 ---SSPERTVPOD 540
Db 649 RSSSSPPPTQTD 662

```

```

Rx MEDLINE-97321807; PubMed-9178509;
Ra Tzermila M., Katsoulou C., Alexandraki D.;
Rt "Sequence analysis of a 33.2 kb segment from the left arm of yeast
Rt chromosome XV reveals eight known genes and ten new open reading
Rt frames including homologues of ABC transporters, inositol
Rt phosphatases and human expressed sequence tags.";
Rt Yeast 13:583-589(1997).
Cc -1- FUNCTION: INVOLVED IN TELOMERIC SILENCING, AS OVEREXPRESSION
Cc RESTORES SILENCING AT HMR IN SIR2 MUTANTS.
Cc -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
Cc -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL: U39041; AAA81033.1; -.
Cc EMBL: L47120; AAB38430.1; -.
Cc EMBL: Z74810; CAA99078.1; -.
Cc SGD: S0005429; HST1.
Cc KX Transcription regulation; Repressor; DNA-binding; Zinc-finger;
Cc Nucleic protein.
Cc FT DOMAIN 200 439 SIR2-TYPE CORE DOMAIN.
Cc ZN_FING 318 345 C4-TYPE (POTENTIAL).
Cc FT SEQUENCE 503 AA; 57702 MW; 4CDF2799E4135ABB CMC64;

```

Query Match 15.4%; Score 592; DB 1; Length 503;
 Best Local Similarity 37.3%; Pred. No. 8.1e-29;
 Matches 138; Conservative 62; Mismatches 104; Indels 66; Gaps 9;

```

Qy 188 TDPR-----TLKDLPEITPPELDMTLMQIYIN-ILSEPPKRRKQDINTEDAVK 240
Db 155 TDPLEKKHAHVLIKDL-----QKAINKVLST--TRLPLPNTIDHFTA 195
Qy 241 LLOECKTIIVLTGAGVSCGIPDRSRDGIYARLAVFPDLDPQAMFDEYFPRDRP 300
Db 196 TLRNAAKTIIVLTGAGVSTSLGIPDRSRSEGFYSK--IRHLGLEDDVDVFNLDIFLDPSV 253
Qy 301 FFKFAKEIYVGOFPDPLCHKFTALSDKEGKLLRNTQNTIDLEQVAGI--QRILOCHGSF 358
Db 254 FYNIAMVLPENMYSPLHSFIKMIQDKKLLRNTQNTIDNLESYAGIDPKLVGCHGSF 313
Qy 359 ATASCLIKIYVDCAVAGDIFNOVPRCPRC-----PAD 393
Db 314 ATASCVTCHWQIPGEKIPENIRNELPLCPYCYOKRKQYFPMSNGNNTVQTNINENSPIL 373
Qy 394 EPLAIMKEIVFEGENLPEQFHRAMKYDKDEVDLLIVIGSSLKVRPALIPSSIPHEVPO 453
Db 374 KSYGVKLPDMTFGEALSRHKTIKIDLECDLLICIGTSLKAVPSEIVNMVSHVPO 433
Qy 454 ILINREPLPHLFDVLLGDCDVYIN-----ELCHRLGGEYAKLCCNPYKLSIET--503
Db 434 ILINDDVYTHAEFDNLGLFCDDVAVSLAKKCHMDIPKKNQDLDKKIDYINCTEIDKGYK 493
Qy 504 -EKPPRPOKE 512
Db 494 IKQPRKKQO 503

```

RESULT 2
 ID HST1_YEAST STANDARD: PRT; 503 AA.
 AC P53685;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST1 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 1).
 GN HST1 OR YOL068C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetales; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / YPH1;
 RX MEDLINE-96101589; PubMed-7498786;
 RA Brechmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RA Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GRT 668;
 RX MEDLINE-96405912; PubMed-8810037;
 RA Debyshire M.K., Weinstock K.G., Strathern J.N.;
 RT "HST1, a new member of the SIR2 family of genes.";
 RL Yeast 12:631-640(1996).
 RN [3]
 RP SEQUENCE FROM N.A.

RESULT 3
 ID SIR2_CANAL STANDARD: PRT; 515 AA.
 AC O59923;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE REGULATORY PROTEIN SIR2.
 GN SIR2.

OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC anamorphic Saccharomycetales; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martin J., Johnson A.D.;
 RT "SIR2 gene from Candida albicans."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF045774; AAC09304.1; -;
 KM Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN_FING 352 453 SIR2-TYPE CORE DOMAIN.
 FT C4_FING 379 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 515 AA; 57732 MW; 2E4DB259EFA9251F CRC64;

 Query Match 14.8%; Score 568.5; DB 1; Length 515;
 Best Local Similarity 39.4%; Pred. No. 2.2e-27;
 Matches 141; Conservative 60; Mismatches 122; Indels 35; Gaps 10;

 QY 138 LILTDGLTNGFHSCESD--DDRTSHA---SSDWTFRRIQGYTVOQHLMGTDPRT 192
 DB 141 LMLTDSNQLPEPKVEIDENDKNDGTNSDIDSNDMSQSESG-----ELGDAMD-- 190
 QY 193 ILKDLLETIPPELD-DMTLMOIVINILSEPPKRRKRDINTEDAVKLLQECKTIIVL 251
 DB 191 -VDSLPEP--EDDYDDMSTTLKRTINMTPEKRYK-----LPDLISLSRAKKIMVY 240
 QY 252 TGAGVSVSGIPDFRSRGIYARLAVDPDLPQAMFIEFRKDRPPEKFAKEIYPG 311
 DB 241 TGAGISTSLGIPDFRSRFGIYNLSK--LNLSDPQKVFEDQTFWRREGRLFYTAHLVLP 298
 QY 312 QPSPSLCHKRIALSDEKGLLRNTONTIDPLEOVAGI--QRILQCHSFPATASCLTCKYK 369
 DB 299 DGFESLHAFLKLLQDHNKLRNTONIDNLEQRAKSEKLVQCHSFAKAKVSCQGI 358
 QY 370 VDEEAVAGDIFNOVPRCPRC-----PADPELAIMKPEIYFEGENLEPQGFRAKVKYKDE 424
 DB 359 FAEKRTYNHTRKROVPCALCKNTKQAPRHFGAIKRTITTFEGEDLPERHHTLMDKRLQ 418
 QY 425 VDLIIYGSSLAKVPVALIPSSIPHEVPOQLINREPLPHLHFDVLLGDCDVIINELC 482
 DB 419 IDFLVLIGTSLKVEPVASIIERYVKPKLILNKDPIPNNGFNQLGLGDDNAVSYLC 476

 RESULT 4
 SIR2_YEAST
 ID SIR2_YEAST STANDARD; PRT; 562 AA.
 AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 GN SIR2 OR MAR1 OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85126876; PubMed=6098447;
 RA Shore D., Squire M., Nasmyth K.A.;
 RT "Characterization of two genes required for the position-effect
 RT control of yeast mating-type genes."
 RL EMBO J. 3:2817-2823(1984).
 CC [2]
 CC SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Paulin L., Saren A.M., Laamanen P.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X01419; CAA25667.1; -;
 DR EMBL: Z71781; CAA96447.1; -;
 DR EMBL: Z74090; CAA98600.1; -;
 DR PIR: S05891; KGBYS2.
 DR SCD: S0002200; SIR2.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT ZN_FING 372 498 SIR2-TYPE CORE DOMAIN.
 FT C4_FING 399 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 562 AA; 63261 MW; 52E693753654586 CRC64;

 Query Match 14.7%; Score 568; DB 1; Length 562;
 Best Local Similarity 39.1%; Pred. No. 2.6e-27;
 Matches 127; Conservative 48; Mismatches 92; Indels 58; Gaps 7;

 QY 216 VINILSEPPKRRK---KDIN-----TIEDAVKLLQECKTIIVLG 253
 DB 205 VTINV--EDPLAKQTVRLINDQRAINKVLCIRLSNFTIDHFIQKLTARKILVLTG 262
 QY 254 AGVSVSGGIDPFNSRSDGIYARLAVDPDLPQAMFIEFRKDRPPEKFAKEIYQGF 313
 DB 263 AGVSTSLGIDPFNSSEFYSK--IKHLGLDDPOVFNYNFMHDPSEFYVIANNVLPPEK 320
 QY 314 QPSPSLCHKRIALSDEKGLLRNTONTIDPLEOVAGI--QRILQCHSFPATASCLTCKYKVD 371
 DB 321 TYSPLHSFIMLQMKGGLLRNTONIDNLSYAGISTDKLVQCHGSPATATCVCHNNLP 380
 QY 372 CEAVRGDIFNOVPRCPRC-----PADPELAIMKPEIYFEGENLEPQGFRAKVKYKDE 424
 DB 381 GERIFKNIRNDELPLCYCKKRRYFPEGYNKVVGAASQSSSEKPPYILNSYGLKRP 440
 QY 402 EIVFEGENLPEQFHAKMYKDEVDLIIYGSSLAKVPVALIPSSIPHEVPOQLINREPL 461
 DB 441 DITFEFGALPNKFKSITREDILECDLICIGTSLKVAVSEIVMVPSPQVYLINRDPV 500
 QY 462 PHLHFDVLLGDCDVIINELCHRLG 486
 DB 501 KHAFFDLSTLGYCDDIAMVAQKCG 525

 RESULT 5
 SIR2_LEIMA
 ID SIR2_LEIMA STANDARD; PRT; 381 AA.
 AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 GN SIR2 OR MAR1 OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

AC Q25337; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSTR2RP).
GN SIR2.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]
RP Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
RX MEDLINE-96186914; PubMed-8635734;
RA Yahaoui B., Talbi A., Ouassil A.;
RT "A Leishmania major protein with extensive homology to silent
information regulator 2 of Saccharomyces cerevisiae."
RL Gene 169:115-118(1996).
CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, LA0331; AB06804.1; -
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 148 151 POLY-ALA.
FT ZN-FING 152 179 C4-TYPE (POTENTIAL).
FT DOMAIN 271 278 POLY-SER.
FT DOMAIN 305 310 POLY-SER.
SQ SEQUENCE 361 AA; 41958 MM; 5A311630AA5D2365 CRC64;

Query Match 12.6%; Score 484; DB 1; Length 381;
Best Local Similarity 32.1%; Pred. No. 1.9e-22;
Matches 125; Conservative 55; Mismatches 101; Indels 108; Gaps 12;

QY 220 LSEPPRRKKRKDINTIEDAVKLLQD--CKIIVTLGAGVSCGIPDRSRD-GIYARLA 276
DB 14 LGEF-----VEGLARIYREKDVRLIVLVAGASVAAGIPDRSSPTGIYAKLG 63

QY 277 VDFPDLPPQAMFIEYRKDRPFKFAKE--IYGPQPSLCHKFTALSDKSKLLRN 334
DB 64 K-VYLDPTAFSLTLREKREIFYSIARELNLMPGHOPFAVHNHFTRLQDEGRLLRC 121

QY 335 YTONIDLEOVAGI--OHILCOHGFATASCILCKYKDCVNRDITNOVPRCPRA 392
DB 122 CTQNDIGLEKAGVSPDLVEAHGSAACCECHTPSIEQNYLEAMSGVSRCTGC- 180

QY 393 DEPLAIKPEIYFEGENLPEQFHRAMKXDKDEVLLIYSSILKVRPALIPSSISHP 452
DB 181 ----GIVAPNVYFEGENLPEAFDHALHNDAPIALVELIITIGSMQVHPALLPCVVPKSLP 236

QY 453 QILNRE-----PLPHLH----- 465
DB 237 RVLNMRHVGLLFRPDDPLDTTIDDAVAKESRSSSSQSPASARREGTEGSSS 296

QY 466 -----FDVELLGDGVINILCHRLG--GEYATLC 493
DB 297 PNEVEDASTSSSDGYQYGDYVYAHDPVCRDVFERRGDCQENVLKLACGLREALARRC 356

QY 494 CNPVLSITEKPRPQKELVHLSLPT 522
DB 357 ASPVR-HQLRHRMRMRRE-----SELLPT 380

RESULT 6
HST2_YEAST STANDARD; PRT; 357 AA.
AC P53686;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE HST2 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 2).
GN HST2 OR YPL015C OR LPAC.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
in silencing, cell cycle progression, and chromosome stability."
RL Genes Dev. 9:2888-2902(1995).
CC [2]
RP SEQUENCE FROM N.A.
RA Hall J., Ahmed A., Bussey H., Fortin N., Friesen J.D., Storms R.K.,
Vo D.H., Wang Y., Winnett E.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD FUNCTION IN TELOMERIC SILENCING, CELL CYCLE
PROGRESSION AND CHROMOSOME STABILITY.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, U39063; AAA81035.1; -
DR EMBL, U39335; AB68090.1; -
DR SGD: S0005936; HST2.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 22 250 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 143 173 C4-TYPE (POTENTIAL).
SQ SEQUENCE 357 AA; 39979 MM; ED281E5B8241A4D0 CRC64;

Query Match 11.4%; Score 439; DB 1; Length 357;
Best Local Similarity 32.7%; Pred. No. 9.1e-20;
Matches 117; Conservative 58; Mismatches 119; Indels 64; Gaps 11;

QY 247 KIIVLTGAGVSCGIPDRS-RDGIYARLAVDFPDLPPQAMFIEYRKDRPFKFA 305
DB 26 KIVFVWAGISTSCGIPDRSPGTGLYHNLA--RLKLYPEAVFVDFPDSPLPEYTLA 83

QY 306 KEIYGFQPSLCHKFTALSDKSKLLRNNTQNTDLEOVAGI--RILOCHGFATASC 363
DB 84 KELYGNFRPSFHTLLLFQDKDYALKRYVTONIDLEKQGVKQDLITIAHGSAFHCHC 143

QY 364 LICKYKVCVCEAVNDFIENQVPRCPRADEPLAIKPEIYFEGENLPEQFHRAMKXDK 423
DB 144 IGGCVVPRPYQFKSLAHPIKDPKCC--DVCGELVRAIYFEGEDLDPSSEFWLNSE 201

QY 424 -----EVDLLIYIGSSILKVRPALIPSSISHPQIILNREPLPHLHDEL 470
DB 202 WLREKITTSGRHPQPLIYVGTSLAVYPFASLPEIRKAYKRVLCN-----LET 251

QY 471 LGD-----QDVIINELCHRLGGEYATLCNPNVLSITEKPRPQ-----KE 512
DB 252 VGDFAKMRPTDLYV---HOTSDEFAQLVEELGQWQDFEKLITLAQCGMDNSKEOLLE 307


```

QY 513 LVHLSELPPPLHISEDSSPERTVPDSSVIATLVQATNNVNDLEVSSESCVEEK 570
DB 308 IVDHLE-----NLSDQSEHSADKKDKKL-----QRLNGHDSDDGASNSSSQK 353

RESULT 7
HST4_YEAST STANDARD; PRT; 447 AA.
ID HST4_YEAST
AC P53687;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
GN HST4 OR YOR025W OR OR26.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1;
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RA de Haan M., Maarse A.C., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
CC AND GENOMIC MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39062; AAA81034.1; -
DR EMBL: X87331; CAA60741.1; -
DR EMBL: Z74933; CAA39215.1; -
DR SGD: S000551; HST3.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 50 314 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 195 223 C4-TYPE (POTENTIAL).
SQ SEQUENCE 447 AA; 50524 MW; 59CD5FBD712B7005 CRC64;

Query Match 8.9%; Score 341.5; DB 1; Length 447;
Best Local Similarity 28.3%; Pred. No. 9,4e-14;
Matches 119; Conservative 60; Mismatches 147; Indels 95; Gaps 19;

QY 231 DINTIEDAVALLQCKKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVD---FPDLPPQAA 287
DB 38 DDEVLLRIVTQQLSRRIACLTGAGISCMAGIDPFRSSDGLYLVKDCQOYMSIKSGRE 97
QY 288 MEDIEYFRKQPR--PEPFKAKEIYPG--QFQSLCHKFIALLSDEKGLRNTQNTDITL 343
DB 98 MEDISLFRDQFKISIFAKFEMERLYSNVQLAKPFTKHFIAHLKDRNKLRLCYTONIDGLE 157
QY 344 QVAGIQ-----RIIQGSGFATASCLC-----KKVDCNAVGRD 378
DB 158 ESIGLTLNKRKLPLTSSSSHMKNLDVVLHGDKLTISCTCFQTFPMRSYWSRC-LRGE 216
QY 379 IFNQVVRPCRPCPA-----DEPLAIKKPEIVFEGENLP--EQFHRAKMD-- 421

```

```

DB 217 L-----PLCPDCEALINKRLNEGKRTLGSNVGLIRPVIYLYGENHSPCELLITGGLNDII 271
QY 422 KDEVDLLIVTGSSLKXAPVALLIPSSIPHEVQ-----ILINRELP---HLHFDEYLLG 472
DB 272 KGNPDFLITMGTSLKVDGVQLVKLSKKIHDRGLIILVKNKPTIGSSWHGJIIDVQIHS 331
QY 473 DCOYIINELCHRIIGGEVAKICCNPVKLSI-----TEKPPQKELVHLSELPPPLHI 526
DB 332 DCDMMWTFLESQI-PDEFKIQDDIKRLQLKREASDLRKQKQKQDSI---GTPP----- 383
QY 527 SEDSSPERTVPDSSVIATLVQATNNVNDLEVSSESCVEEK---POEVQTSRNVENI 583
DB 384 -----FPLRT-----AGIDIGGNNEIN-TIKSLNLYVKRKILSPENSSEDEBENL 429
QY 584 N 584
DB 430 D 430

RESULT 8
HST4_YEAST STANDARD; PRT; 370 AA.
ID HST4_YEAST
AC P53688;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 4).
GN HST4 OR YDR191W OR YD9346.03.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP GENE NAME.
RC STRAIN-GRY 668;
RX MEDLINE-96405912; PubMed-8810037;
RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
RT "HST1, a new member of the SIR2 family of genes.";
RL Yeast 12:631-640(1996).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST3 TO TELOMERIC SILENCING,
CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
CC AND GENOMIC MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48784; CAA88705.1; -
DR SGD: S0002599; HST4.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 90 342 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 221 254 C4-TYPE (POTENTIAL).
SQ SEQUENCE 370 AA; 41765 MW; 86BB0238BF914F1 CRC64;

```



```

RX MEDLINE-90150286; PubMed-1689217;
RA Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT "CNS stem cells express a new class of intermediate filament
RT protein."
RL Cell 60:585-595(1990).
CC -1- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -1- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M34384; AAA41685.1; -.
DR PIR: A34736; A34736.
DR INTERPRO: IPR001664; -.
DR PFAM: PF000038; filament; 2.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Heptad repeat pattern;
KW Neurone.
FT DOMAIN 1 7 HEAD.
FT 314 314 ROD.
FT 1805 1805 TAIL.
FT 8 43 COIL 1A.
FT 44 55 LINKER 1.
FT 56 151 COIL 1B.
FT 152 174 LINKER 12.
FT 175 193 COIL 2A.
FT 194 196 LINKER 2.
FT 197 314 COIL 2A.
FT 197 314 COIL 2A.
SQ SEQUENCE 1805 AA; 198744 MW; BA0EE14717E0998D CRC64;

Query Match 3.5%; Score 135; DB 1; Length 1805;
Best Local Similarity 19.7%; Pred. No. 1.7;
Matches 155; Conservative 99; Mismatches 262; Indels 272; Gaps 36;

QY 41 PGGRSGPGEAANVAPAAACCEAASAAAPALMREAAGAANA---ERAPA----- 89
DB 392 PPISEAPCPPEANAEVRAQOEVPFLSLQTOAPPEPLMKATVPSSSAIPLEEGGKQGHF 451
QY 90 -----TAVAGDGDNGSGSLRREPRADDFDDEGE-----EEDENAAAA 127
DB 452 DDLTSLATNLNPHHPTLEAKDGESS-----ESRVSSIFQDEQIWEIVKEADIEVKE 506
QY 128 AAAATGVRNLLITDGLTNGFHSCESDDDRISHASSSWTPRRPGPTTFVOQHLM-- 185
DB 507 NSSAQKTOESGLDTE-----ETQDSQ-----GP---LQKETIKA 537
QY 186 IGTDPRTILKDLPEITPPPELDQMTLQIVINILSEPPRK-----RKRD---NTI 235
DB 538 LGSEPLMSLKTQVETAGKNCNSST--EGHLGLGSEPEKEKQIPKLSLEKNVESKTL 595
QY 236 EDVAVKLLQECKKIIIVLGAGVSVSGIPDRSRDGIYARLAVDFPDLDPQAFMEDIEYFR 295
DB 596 ENGVPVLSL-----LQKEDRTD-----QELMSPKGTL----- 625
QY 296 KDRPPFFKAKETIYPCGFOPSLCHKFTALSDKGLKIRNTQNTIDTLEQVAGIQRILQCH 355
DB 626 ---KRSSSLKE-----SOEVRPSKSGN--LESWTAKKEESQHLG----- 662
QY 356 GSFATSCCLCKYKVDCEAVRGDIENQVVRPCRPADDELALMKPPIVFGENTLPQGH 415
DB 663 --FPGAEQMLERLVEKED-----QSPRSPSEEDQECR-----LQENQ 702
QY 416 RANKYDKDE---VDLIVIGSSLKVRVALIPSSIPHEVPOIL--INREPLPHLHEVEL 470
DB 703 EPIGYEBAEQILIERLIKESQESLRS---PEEEDQEAQRSLOKENGDEPLGYEAE--- 755

```

```

QY 471 LGDDVILNELCHRLGGEYAKLCNPKYLSITEKPPRP-QKELVHLSLPP----- 522
DB 756 ---DQMLERLIKESQESLK---SPENORIGPLPERNQKSLRYLEENQETFPVLESR 808
QY 523 -----PLH-----ISDSSSPERTVQDSVATLVDAQTN 554
DB 809 NQRLRLSEVEEERQIRYKLEKVSQDSLSLAENQPLRYLEEDCIKSLIEDKTHK 868
QY 555 NVNMLE-----VSSSCVEEKPOEVQTSRVNENIVNENPDKRAGSSADKNER 603
DB 869 SLGLEDNRNDSTIIIPQESFTQ-VSLRPPEEDQRIYVNHLEKESQEF-----SR 916
QY 604 TSAVEIYRKCPNPLNLAEGQSKR-LEGNQILFVPPNRYTHGAEVYSDSDVLSSSCG 662
DB 917 SSEE-----EQVMERSLEGEN-----HESLSVYKEKEDQWVESQLEK 953
QY 663 SNSDGTGQSPSLPEPLEDSEIEFYNGLEDTEREPRECAAGSGFGADGGQEVVNEAIA 722
DB 954 ESQDSG-----KSLDEDS--QETGPLEKENAE-----SLRSLAGQDQEEQKLEQ 997
QY 723 TRQELTDV 730
DB 998 TQOTRLRAY 1005

RESULT 11
KMLS_CHICK STANDARD; PRT; 1906 AA.
AC P11799; P19038;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES
DE (EC 2.7.1.117) (MCK) [CONTAINS: TELOKIN].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A. (MCK-210).
RX MEDLINE-96033976; PubMed-7589469;
RA Watterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,
RA Stepanova O.V., Shtrinsky V.P.;
RT "Multiple gene products are produced from a novel protein kinase
RT transcription region."
RL FEBS Lett. 373:217-220(1995).
RN [2]
RP SEQUENCE FROM N.A. (MCK-108).
RX MEDLINE-90192792; PubMed-2315320;
RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
RA Means A.R.;
RT "Regulatory and structural motifs of chicken gizzard myosin light
RT chain kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
RN [3]
RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE-FIBROBLAST;
RC MEDLINE-90361738; PubMed-2202734;
RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
RA van Eldik L.J., Watterson D.M.;
RT "Use of DNA sequence and mutant analyses and antisense
RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
RT myosin light chain kinase autoinhibition, calmodulin recognition, and
RT activity."
RL J. Cell Biol. 111:1107-1125(1990).
RN [4]
RP SEQUENCE OF 1259-1906 FROM N.A.
RC TISSUE-GIZZARD;
RX MEDLINE-87157587; PubMed-3030394;
RA Guerlero V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;
RT "Domain organization of chicken gizzard myosin light chain kinase

```

RT deduced from a cloned cDNA.";
 RL Biochemistry 25:8372-8381(1986).
 RN [5]
 RN SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RP TISSUE-GIZZARD:
 RC MEDLINE-93073972; PubMed-1444462;
 RX Yoshikai S., Ikebe M.;
 RA "Molecular cloning of the chicken gizzard telokin gene and cDNA";
 RL Arch. Biochem. Biophys. 299:242-247(1992).
 RL [6]
 RP SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RA MEDLINE-92236611; PubMed-1373815;
 RA Collinge M., Matristian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
 RA van Eldik L.J., Watterson D.M.;
 RT "Structure and expression of a calcium-binding protein gene contained
 within a calmodulin-regulated protein kinase gene.";
 RL Mol. Cell. Biol. 12:2359-2371(1992)
 CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A
 MYOSIN LIGHT CHAIN, WHICH LEADS TO THE FORMATION CALMODULIN/
 TRANSDUCTION OF CALCIUM SIGNALS.
 CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN LIGHT-CHAIN] - ADP + [MYOSIN
 LIGHT-CHAIN] PHOSPHATE.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS OF MLCK ARE PRODUCED
 BY ALTERNATIVE INITIATION, A NON-MUSCLE FORM (MLCK-210; WHICH IS
 THE LONGEST FORM); A SMOOTH-MUSCLE FORM (MLCK-108) AND TELOKIN (A
 C-TERMINAL SECTION WITH NO CATALYTIC ACTIVITY).
 CC -1- TISSUE SPECIFICITY: TELOKIN IS EXPRESSED IN GIZZARD, HEART, LUNG,
 INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF THE
 EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE GIZZARD.
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X52876; CA37056.1; -;
 DR EMBL; X52876; CA37057.1; -;
 DR EMBL; X52876; CA37058.1; -;
 DR EMBL; M31048; AAA49069.1; -;
 DR EMBL; M14953; AAA69964.1; -;
 DR EMBL; M96655; AAA49083.1; -;
 DR EMBL; M88283; AAA48647.1; -;
 DR EMBL; M88284; AAB53768.1; -;
 DR PIR; A25810; A25810. -;
 DR PIR; A37099; A37099. -;
 DR PIR; S11652; S11652. -;
 DR INTERPRO: IPR000719; -;
 DR INTERPRO: IPR001777; -;
 DR INTERPRO: IPR002290; -;
 DR INTERPRO: IPR003006; -;
 DR PFAM: PF00041; fn3.1. -;
 DR PFAM: PF00047; Ig.9. -;
 DR PFAM: PF00069; pkinase.1. -;
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1. -;
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1. -;
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1. -;
 KW Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
 KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 KW Alternative initiation.
 FT CHAIN 1 1906 MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
 FT CHAIN 935 1906 MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
 FT CHAIN 1750 1906 ISOZYME (MLCK-210).
 FT CHAIN 1750 1906 ISOZYME (MLCK-108).
 FT CHAIN 1750 1906 TELOKIN.

FT INIT_MET 935 935 FOR MLCK-108.
 FT INIT_MET 1750 1750 FOR TELOKIN.
 FT DOMAIN 1330 1400 FIBRONECTIN TYPE-III.
 FT DOMAIN 1453 1708 PROTEIN KINASE.
 FT DOMAIN 1809 1976 IG-LIKE C2-TYPE DOMAIN.
 FT NP_BIND 1459 1467 ATP (BY SIMILARITY).
 FT BINDING 1482 1482 ATP (BY SIMILARITY).
 FT ACT_SITE 1574 1574 BY SIMILARITY.
 FT DOMAIN 1716 1728 CALMODULIN AUTOINHIBITION (AM13) REGION
 FT DOMAIN 1730 1749 CALMODULIN RECOGNITION (RS20) REGION
 FT DOMAIN 1317 1364 (POTENTIAL).
 FT DOMAIN 1385 1402 MOTIF IA.
 FT DOMAIN 660 1833 MOTIF IB.
 FT REPEAT 660 676 4 X REPEATS, MOTIF IIA.
 FT REPEAT 758 774 IIA-1.
 FT REPEAT 1107 1123 IIA-2.
 FT REPEAT 1817 1833 IIA-3.
 FT DOMAIN 693 1866 IIA-4.
 FT REPEAT 693 708 5 X REPEATS, MOTIF IIB.
 FT REPEAT 791 807 IIB-1.
 FT REPEAT 1140 1156 IIB-2.
 FT REPEAT 1281 1297 IIB-3.
 FT REPEAT 1851 1866 IIB-4.
 FT DOMAIN 970 1226 IIB-5.
 FT REPEAT 970 987 4 X REPEATS, MOTIF III.
 FT REPEAT 999 1016 III-1.
 FT REPEAT 1061 1078 III-2.
 FT REPEAT 1209 1226 III-3.
 FT DOMAIN 1700 1763 III-4.
 FT DOMAIN 1896 1906 CALMODULIN-BINDING.
 FT MOD_RES 1748 1748 POLY-GLU.
 FT MOD_RES 1762 1762 PHOSPHORYLATION.
 FT CONFLICT 1439 1439 PHOSPHORYLATION.
 FT SEQUENCE 1906 AA: 210445 MM; AD/D8A3B65EE3363 CRC64;
 SO
 Query Match 3.5%; Score 133; DB 1; Length 1906;
 Best Local Similarity 19.0%; Pred. No. 2.5;
 Matches 150; Conservative 95; Mismatches 257; Indels 288; Gaps 35;
 QY 65 SAAPALMREAAAGAAASAEAPATAVAGDGN-----GSGLR--EPRADDFD 113
 DB 827 SSASRAEMLRDGRBSASGGER-----DGNYGALFTGRTSGFKSSSTRABEEQ 878
 QY 114 DD-----EGEEDDAAAAAAAAAIGYRDNLITDGLITNFHGCESDDRTSHA 163
 DB 879 EDVGVGLKRVETREHTEESLRQDEQLDFRD--ILGKVKSTKSF---SEEDLKEIPA 932
 QY 164 SSDWTP--RPRIGPYTFVQ---HLMIGTDPRTIL-----KDLPETIPPELDDMT 211
 DB 933 EQMDFRANLQGVKKTITSEERKYNHARQOYDFRVLAKKGTPKTPLEKVPKRP-PAVT 991
 QY 212 LMQIVINILSEPPKRRKKRKDDINTJEDAVKLLQECKKIIVLGAGVSVSCGIPDFRSRDCI 271
 DB 992 DFRSVLGAKKRPAB-----NGSASTPAPNARAGS 1021
 QY 272 YARLAVDFPDLDPDQAMEDIYFRKDPFRFAKEIYPGQFSPSLCKFIATLSKEKL 331
 DB 1022 EAQNTAPNSEAP-----APRPVK-----KEEKN 1045
 QY 332 LRNYQNIDTLEQVAGIQLQC--HGSFATASCLCKKYKVCDAVGRDIFNOVPRCPRC 390
 DB 1046 DR-----KCEHG-----CAYVDGIIIGKKAENKPA 1071
 QY 391 PADEPLAMKEIYFEGENLPEQFHRAMKYDKDEVLLIVIGSSLKVRPAVALPSSIPHE 450
 DB 1072 SKPTP-----PPSKTAPSFTEKLQDAKAVDGEKVLGCRISSDPPASVMTLDSKAITS 1126
 QY 451 VPQILINPEPLPHLFDVLLGD--CDVIINELCHRLGSEY-----AKLC----- 494
 DB 1127 SKSIVISO-----GLTCLSTIIEKVMPEDEGGEYKCIENAGACACAKVAVE 1174

RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96290553; PubMed-8703835;
 RA Nilsson I., Loechner K., Siegler G., Grell J., Beck J.D., Fey G.H.,
 RS Marschalek R.;
 RT "Exon/Intron structure of the human ALL-1 (MLL) gene involved in
 RL translocations on chromosomal region 11q23 and acute leukaemias.";
 RN Br. J. Haematol. 93:966-972(1996).
 [3]
 RP SEQUENCE OF 1-1909 FROM N.A.
 RX MEDLINE-93390935; PubMed-8378076;
 RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
 RS Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
 RT "Two distinct portions of Irf1/ENL at 19p13 are involved in t(11;19)
 RL leukemia.";
 RN Oncogene 8:2617-2625(1993).
 [4]
 RP SEQUENCE OF 1317-2328 FROM N.A.
 RX TISSUE-BRAIN;
 RS MEDLINE-93265134; PubMed-1303259;
 RA Djabbal M., Seller L., Parry P., Bower M., Young B.D., Evans G.A.;
 RT "A trithorax-like gene is interrupted by chromosome 11q23
 RL translocations in acute leukaemias.";
 RN Nat. Genet. 2:113-118(1992).
 [5]
 RP SEQUENCE OF 1251-1538 FROM N.A.
 RX MEDLINE-94215165; PubMed-8162575;
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaan O.,
 RS Saito H., Croce C.M., Canaan E.;
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 RL involved in acute leukaemia.";
 RN Cancer Res. 54:2326-2330(1994).
 [6]
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RX MEDLINE-95322025; PubMed-7598802;
 RA Mbangkollo D., Burnett R., McCabe N., Thirman M., Gall H., Yu H.,
 RS Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RL trx zinc-finger domain, and alternative splicing.";
 RN DNA Cell Biol. 14:475-483(1995).
 [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE-95315013; PubMed-794749;
 RA Marschalek R., Grell J., Loechner K., Nilsson I., Siegler G.,
 RS Zwechbrunner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 RL transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RN translocation t(4;11).";
 RN Br. J. Haematol. 90:308-320(1995).
 [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE-94020842; PubMed-841518;
 RA Forster A., Rabbitts T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 RL application to isolation of MLL fusion cDNAs from acute leukaemia
 translocations.";
 RN Oncogene 8:3157-3160(1993).
 [9]
 RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND ENL, T(4;11)(Q21;Q23)
 CC THAT INVOLVES MLL AND APL, T(9;11)(P22;Q23) THAT INVOLVES MLL AND
 CC APL, T(6;11)(Q27;Q23) THAT INVOLVES MLL AND APL, T(11;17)(Q23;Q21) THAT INVOLVES MLL
 CC AND APL, T(10;11)(P12;Q23) THAT INVOLVES MLL AND APL,
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND ENL AND T(11;11)(Q21;Q23)
 CC THAT INVOLVES MLL AND APL.
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
 CC FACTORS.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04284; AAA58669.1; -
 DR EMBL: 269744; CAA93625.1; -
 DR EMBL: 269745; CAA93625.1; JOINED.
 DR EMBL: 269746; CAA93625.1; JOINED.
 DR EMBL: 269747; CAA93625.1; JOINED.
 DR EMBL: 269748; CAA93625.1; JOINED.
 DR EMBL: 269749; CAA93625.1; JOINED.
 DR EMBL: 269750; CAA93625.1; JOINED.
 DR EMBL: 269751; CAA93625.1; JOINED.
 DR EMBL: 269752; CAA93625.1; JOINED.
 DR EMBL: 269753; CAA93625.1; JOINED.
 DR EMBL: 269754; CAA93625.1; JOINED.
 DR EMBL: 269755; CAA93625.1; JOINED.
 DR EMBL: 269756; CAA93625.1; JOINED.
 DR EMBL: 269757; CAA93625.1; JOINED.
 DR EMBL: 269758; CAA93625.1; JOINED.
 DR EMBL: 269759; CAA93625.1; JOINED.
 DR EMBL: 269760; CAA93625.1; JOINED.
 DR EMBL: 269761; CAA93625.1; JOINED.
 DR EMBL: 269762; CAA93625.1; JOINED.
 DR EMBL: 269763; CAA93625.1; JOINED.
 DR EMBL: 269764; CAA93625.1; JOINED.
 DR EMBL: 269765; CAA93625.1; JOINED.
 DR EMBL: 269766; CAA93625.1; JOINED.
 DR EMBL: 269767; CAA93625.1; JOINED.
 DR EMBL: 269768; CAA93625.1; JOINED.
 DR EMBL: 269769; CAA93625.1; JOINED.
 DR EMBL: 269770; CAA93625.1; JOINED.
 DR EMBL: 269771; CAA93625.1; JOINED.
 DR EMBL: 269772; CAA93625.1; JOINED.
 DR EMBL: 269773; CAA93625.1; JOINED.
 DR EMBL: 269774; CAA93625.1; JOINED.
 DR EMBL: 269775; CAA93625.1; JOINED.
 DR EMBL: 269776; CAA93625.1; JOINED.
 DR EMBL: 269777; CAA93625.1; JOINED.
 DR EMBL: 269778; CAA93625.1; JOINED.
 DR EMBL: 269779; CAA93625.1; JOINED.
 DR EMBL: 269780; CAA93625.1; JOINED.
 DR EMBL: D14540; BAA03407.1; -
 DR EMBL: L01986; AAA92511.1; -
 DR EMBL: D04737; AAA18644.1; -
 DR EMBL: S78570; AAB34770.1; -
 DR EMBL: X83604; CAA58584.1; -
 DR EMBL: S66432; AAB28545.1; -
 DR TRANSFAC: T02337; -
 DR MIM: 159555; -
 DR INTERPRO: IPR001214; -
 DR INTERPRO: IPR001487; -
 DR INTERPRO: IPR001965; -
 DR INTERPRO: IPR002857; -
 DR PFAM: PF00628; PHD; 3.
 DR PFAM: PF00856; SET; 1.
 DR PFAM: PF02008; ZF-CXXC; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR K01000; Chromosomal translocation; DNA-binding; Bromodomain;
 DR K01000; Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 DR K01000; Alternative splicing.
 FT DOMAIN 17 102
 FT DNA_BIND 169 180
 FT DNA_BIND 217 227
 FT DNA_BIND 301 309
 FT DOMAIN 1431 1482
 FT DOMAIN 1484 1533
 FT DOMAIN 1566 1627
 FT DOMAIN 1703 1748
 FT BROMODOMAIN (DIVERGENT).
 FT ALA/GLY/SER-RICH.
 FT A.T HOOK (BY SIMILARITY).
 FT A.T HOOK (BY SIMILARITY).
 FT A.T HOOK (BY SIMILARITY).
 FT PHD 1.
 FT PHD 2.
 FT PHD 3.
 FT BROMODOMAIN (DIVERGENT).

```
FT DOMAIN 3840 3969 SET DOMAIN.
FT DOMAIN 137 143 POLY-GLY.
FT DOMAIN 561 564 POLY-PRO.
FT DOMAIN 568 571 POLY-PRO.
FT VARSPLIC 1407 1444 MISSING (IN ISOFORM 14P-18B).
FT CONFLICT 144 144 E -> ELTQIPCSMRKTHGTHDKTEPFLLAWSCLN
FT CONFLICT 317 379 (IN REF. 2).
FT CONFLICT 317 379 GLINSLEKPOKVRKDEGTPPTKEKTYVROSPRRIRK
FT CONFLICT 317 379 VRIIPSSKRTDATTAKOLORA -> VSSLINKMSPKSG
FT CONFLICT 317 379 KTRREHLKOKIROLOSKALGSLQGLFLLQGMQPL
FT CONFLICT 317 379 SNSYRCQ (IN REF. 1).
FT CONFLICT 556 556 Q -> E (IN REF. 2).
FT CONFLICT 1487 1487 R -> G (IN REF. 5).
FT CONFLICT 1603 1603 S -> SGTG (IN REF. 2 AND 7).
FT CONFLICT 1616 1616 O -> H (IN REF. 6).
FT CONFLICT 1937 1937 P -> S (IN REF. 4).
FT CONFLICT 2181 2181 R -> G (IN REF. 2).
FT CONFLICT 3718 3718 N -> D (IN REF. 2).
FT CONFLICT 3759 3759 D -> G (IN REF. 2).
FT CONFLICT 3813 3813 R -> A (IN REF. 2).
FT CONFLICT 3901 3901 R -> A (IN REF. 2).
SO SEQUENCE 3969 AA: 431882 MW: C609CAF082E36CC9 CRC64;
```

Query Match 3.2%; Score 124.5; DB 1; Length 3969;
Best Local Similarity 19.3%; Pred. No. 21;
Matches 130; Conservative 92; Mismatches 234; Indels 217; Gaps 28;

```
QY 36 PRDQPGIGRSGERSAANAACE-----AASAAAPALMREAAGAAAREAP 88
DB 44 PPGGGGGGAPPPPAVAAAAAGSGAGVPGAAASASSSSSSSSSSSSSASGP 103
QY 89 ATAVAGDDN-----GSGLRREPRADDEDEGEDEEAASAAAAAIGRDML 139
DB 104 ALKRGPFDALQVSAIGTLRFRVAFEGSGGGSGGSGDEQ----- 147
QY 140 LTDGLLTNGFHSCEDDRT---SHAASSDWTPR-PRIGYTFVQOHLMTGTPRT-IL 194
DB 148 -----FLGFGSDEVRVRSPTSPSVKTPRGRPRS-----GSDRNSAL 190
QY 195 KDLPETIPPELDMTMOYVINILSEPPKRRKDJNTIEDAVKLQECKKIIVLTGA 254
DB 191 SD-PSVSPSPNKSE-----TKSGCKIKKKDKSIE-----KKRRPPTFPFGV 231
QY 255 GVSVCGIPDFRSRGIYARLAVDEPDDQAMDIYFRKDPFRKFA---KEIYG 311
DB 232 KIKITHG-----KDISELPGKKEDSLKIKKTPATFOQATIKKLRAG 276
QY 312 QFQPSLCHKFTALSDKEGKLLRNTYQNTIDLEQVAGIORIIOCHGSFATASCLICKYVD 371
DB 277 KLSP-LKSKF-----KTGKL-----QIG 293
QY 372 CEAVAGDIFNQVPRCPRADEPLAKPELVFEGENLPEQFHAKMYKD-----EV 425
DB 294 RKGV-----QIVRRGRPPSTE-RIKTPSGLLINSELEP--QKVRKDEGTPPLTKE 343
QY 426 DLLVIGSLKVPALIPSSIPHEVQILINREPLPHIHFEVLLGDCDVITNELCHR- 484
DB 344 DKTIVROSPRRIKPVRIITPSSKRTDA-----TIKKLLORA 379
QY 485 -----LGEYAKLICQNVK-----LSEITKPPRPQKELVHLSELPPPL 524
DB 380 KKAQOKIEKEAALQGRKVKVQYKINROFIMPVYSAISSRIKPRPFIEDEDYD-P-PI 438
QY 525 HISESSSPER--TYPQ--DSSVIALVQATRNANNVNDLEVSSESCVEEKPQEVOTSRV 581
DB 439 KIRLESTPNRSFAPSGSSSEKSSAASQHSOMSSDSRSSPSVDYS--TDSQASEEIQ 497
QY 582 NINVENPEKAV-----GSSADKKNERTSVAE-----TVRKCPWRRLAKELIS 624
DB 498 VLPERSRTPPEYHPPPLISQSPENESNDRSRKRYSVSRSGSRTTKLSTLOSAPQOOT 557
QY 625 KRLEGNQYLFVPR 637
```

```
DB 558 SSSPPPLTPPP 570
RESULT 15
ICP4_HSVB STANDARD; PRT: 1487 AA.
ID ICP4_HSVB
AC P28925;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4 (155 KDA IMMEDIATE-EARLY
DE PROTEIN).
DE IE OR 64.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -I- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -I- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -I- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -I- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M86664; AB02515.1; -
DR EMBL; M86664; AB02499.1; -
DR PIR; A36802; EDEBF6.
DR HSSP; P04002; 1WFA.
DR TRANSFAC; T00401; -
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 181 213 SER-RICH.
FT DOMAIN 922 931 ARG/LYS-RICH (BASIC).
SO SEQUENCE 1487 AA: 154868 MW: 1D439783BD03680D CRC64;
```

Query Match 3.2%; Score 124; DB 1; Length 1487;
Best Local Similarity 25.7%; Pred. No. 6.3;
Matches 45; Conservative 19; Mismatches 71; Indels 40; Gaps 5;

```
QY 8 ALQAGSPSAAMAEAAQAPDEPLRKRRPRRDGPGIGR-----SPGEPSAANAPAA 58
DB 87 AAGSGPPTPAAPAS-PAPSP-----APLALMLKRVHSSVAPGKRRRTGSSS 137
QY 59 AGCEAASAPAAALWREAAGAAASAREAPATAVAGDDNGSLRREPRADDFDDECE 118
DB 138 PG--CGDAADPVALDSDETFCPSQPEFPSSASPGGSPAPRVRSISISSSSSSSMD 195
QY 119 EDEEAASAAAAAIGRDNLLITDGLLTNGHSCSDDDDRSHASSSDMTTPRR 173
DB 196 EDDQADGAGASS-----SSSSSDSDSDSGGGEETPRPR 230
```

Search completed: February 16, 2001, 10:52:51
Job time: 554 sec

Tue Feb 27 15:47:14 2001

us-09-461-580a-26.rsp

Page 13

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 16, 2001, 10:44:13 ; Search time 51.38 Seconds
(without alignments)
257.578 Million cell updates/sec

Title: US-09-461-580A-26

Perfect score: 3854

Sequence: 1 MADEVALALAAAGSPSAAA.....NEATATRGELDVNPSPDKS 737

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/CTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135	3.5	1805	1 US-07-853-913-2	Sequence 2, Appl1
2	122	3.2	738	3 US-08-864-038A-3	Sequence 3, Appl1
3	121	3.1	595	1 US-08-425-069-4	Sequence 4, Appl1
4	121	3.1	595	2 US-08-317-844B-4	Sequence 4, Appl1
5	120	3.1	873	3 US-08-990-140-2	Sequence 2, Appl1
6	119	3.1	180	5 5273901-7	Patent No. 5273901
7	119	3.1	180	5 5482709-6	Patent No. 5482709
8	114	3.0	866	3 US-08-651-999A-1	Sequence 1, Appl1
9	113	2.9	335	2 US-08-405-175A-6	Sequence 6, Appl1
10	112.5	2.9	332	2 US-08-405-175A-5	Sequence 5, Appl1
11	112	2.9	594	3 US-08-729-416C-17	Sequence 17, Appl1
12	111.5	2.9	832	1 US-08-209-747-2	Sequence 2, Appl1
13	111.5	2.9	832	1 US-08-458-298-2	Sequence 2, Appl1
14	108	2.8	98	2 US-07-814-220-2	Sequence 2, Appl1
15	108	2.8	98	2 US-07-812-421-2	Sequence 2, Appl1
16	108	2.8	783	5 5231168-2	Patent No. 5231168
17	105.5	2.7	594	3 US-08-729-416C-7	Sequence 7, Appl1
18	105	2.7	309	2 US-08-405-175A-7	Sequence 7, Appl1
19	104.5	2.7	893	1 US-07-977-434-4	Sequence 4, Appl1
20	104.5	2.7	893	1 US-08-458-819-4	Sequence 4, Appl1
21	104.5	2.7	893	4 PCT-US91-07035-4	Sequence 4, Appl1
22	104.5	2.7	1317	3 US-09-083-521-7	Sequence 7, Appl1
23	104	2.7	1435	2 US-08-568-459A-4	Sequence 4, Appl1
24	104	2.7	1435	2 US-08-487-826B-4	Sequence 4, Appl1
25	103	2.7	629	3 US-08-556-419-22	Sequence 22, Appl1
26	103	2.7	629	3 US-08-556-419-23	Sequence 23, Appl1
27	103	2.7	1503	3 US-08-976-255-14	Sequence 14, Appl1
28	103	2.7	2441	1 US-08-194-468-2	Sequence 2, Appl1

29	103	2.7	2441	3 US-08-961-739-2	Sequence 2, Appl1
30	102.5	2.7	309	2 US-08-405-175A-8	Sequence 8, Appl1
31	102.5	2.7	651	1 US-08-431-080-24	Sequence 24, Appl1
32	102.5	2.7	651	2 US-08-938-534-24	Sequence 24, Appl1
33	102.5	2.7	1043	3 US-08-928-361B-30	Sequence 30, Appl1
34	102.5	2.7	1721	3 US-08-700-651-5	Sequence 5, Appl1
35	102.5	2.7	1721	3 US-08-928-361B-6	Sequence 6, Appl1
36	102	2.6	968	3 US-08-651-999A-7	Sequence 7, Appl1
37	102	2.6	1288	2 US-08-690-473-2	Sequence 2, Appl1
38	102	2.6	1780	1 US-08-769-309A-5	Sequence 5, Appl1
39	102	2.6	1780	3 US-08-994-570-5	Sequence 5, Appl1
40	101.5	2.6	2860	2 US-08-826-267-2	Sequence 2, Appl1
41	101	2.6	281	2 US-08-405-175A-9	Sequence 9, Appl1
42	101	2.6	754	3 US-09-214-564A-2	Sequence 2, Appl1
43	100.5	2.6	731	2 US-08-911-364-1	Sequence 1, Appl1
44	100.5	2.6	792	2 US-08-678-039A-40	Sequence 40, Appl1
45	100	2.6	1514	2 US-08-853-310-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853, 913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-461AANA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-853-913-2

Query Match 3.5%; Score 135; DB 1; Length 1805;
 Best Local Similarity 19.7%; Pred. No. 0.0063;
 Matches 155; Conservative 99; Mismatches 262; Indels 272; Gaps 36;

```

QY 41 PGLGRSPGPSANAPAAAGCAASAPALMREAGAAASA---ERBAP-----89
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 392 PPISEAPCPNNAEVAOEVPLSLQTOQAPPEPLMKATVPSSSAILPELEPGKQGHFP 451
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 -----TAVAGDNGSGLRREPRADDFDDEGE-----EEDAAAA 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 452 DDLSTLANLPHPTLEAKDESS-----ESRVSSITQEDGQWELVEKADLVKE 506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 AAAAIGYRDNLITDGLTNGFHCESDDDDRTSHASSSDWTTPRRIPYTFVQOHLM--185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 507 NSSAQKQESGLDE-----ETQDSQ-----GP---LQKETIKA 537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 IGTDPRTLKDLLEPTTPPELDMTLMQIVINILSEPPKKR-----KRKDI---NTI 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 538 LGEEPLMSTKIONYETAKENCNSST--EGHLGTLEGPEKEQPLKSLKEKNVSEKTL 595
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 EDVAVLLOECKIIVLTGAGVSVCIGIDPFRSDGIYARLAVDFDLDPQAMPDIEYFR 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 596 ENGVAIVISEL-----LGKEDTRIED-----QELMSPKGTL-----625
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 KDPREFEFAKEIYPGOFPSLCHKFIALSDREGKLNNTYONIDTLEQVAGIQRILQCH 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 626 ---KRESSLGKE-----SEVVRPSKEGN-LESWTFKRESQHPLG-----662
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 GSFATASCLIKKYKVDCAVNGDIFNQVPRCPRADEPLAIMKEIVFGEMLPEQFH 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 663 --FPAEDQMLERLVEKED-----QSFPRSPDEEDQECRP-----LQKENQ 702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 RAMKYDKDE---VDLLIYIGSSIAKVRPALIPSSIFHEVPOL--INREPLPHLFDVEL 470
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 703 EPLGEEAEGOILERLIKESQESLRS---PEEDQAGASLOKENEOPGLYEAE---755
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 471 LGDCDVIIINEIHLRGLGEYAKLCNPVKLSITEKPPRP-QKEVLHLSLPT-----522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 756 ---DOMLERLIKESQESLK---SPENORIGKPLERENKSLRLYLENGETFVPLESR 808
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 523 -----PLH-----ISEDSSSPERTVPODSSVIATLVQATNN 554
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 809 NORPLRSLVEEEOQRIYKPLEKVSQDSLGLAEENQVPLRLIEDDCINKSLIEDKTHK 868
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 555 NVNDEL-----VSESSCVERKEQEVQTSRNVENINVENEDFAVGSSTDKMER 603
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 869 SLGSLIEDRNGDSIITPOSETO-VSLRPEEDQRIVNHLEKESQEF-----SR 916
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 TSAVETVAKCWPMLAKQJISKR-LEGNQYLFVPPNRYIFHGAEYVSDSEDDVLSSSCG 662
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 917 SSEEF-----EQVMERSLEGEN-----HESLSSVKEKEDQMVSOLEK 953
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 663 SNSDSGTCQSPSLERPLEDEIEEFYNGLEDDETERPCAGSGGAGDGOEVVNEAIA 722
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 954 ESQDSG-----KSLDES--QETFGPLEKENAE-----SLKSLAGQDQDEBKLPQE 997
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 723 TRQELTDV 730
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 998 TQQTILAV 1005
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2

US-08-864-038a-3
 ; Sequence 3, Application US/08864038A
 ; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

TITLE OF INVENTION: NO SAID POLYPEPTIDE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: 812-5 Hirano
 STREET: Isshinden
 CITY: Tsu-city
 STATE: Mie-prefecture
 COUNTRY: JAPAN
 ZIP: 514-01

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: Microsoft windows 95

SOFTWARE: Word Perfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/864,038A

FILING DATE: May 28, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-184459

FILING DATE: 15-July-1996

ATTORNEY/AGENT INFORMATION:

NAME: C. Bruce Hamburg

REGISTRATION NUMBER: 22,389

REFERENCE/DOCKET NUMBER: F-5610

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)986-2340

TELEFAX: (212)953-7733

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 738.

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Pinctada fucata

CELL TYPE: mantle epithelial cell

FEATURE:

NAME/KEY: peptide

LOCATION: from 1 to 738

IDENTIFICATION METHOD: E (by experiment)

US-08-864-038a-3

Query Match 3.2%; Score 122; DB 3; Length 738;
 Best Local Similarity 32.3%; Pred. No. 0.022;
 Matches 61; Conservative 9; Mismatches 65; Indels 54; Gaps 8;

```

QY 5 VALALQANG-----SPSNAAMEASOPADEPLKRRPRDGPGLGR 45
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 481 LAALAAAGAGGGGFGGLGGLGSGSAAAAAASGSGGFRALRRALRQWRGGGS 540
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 46 SPGEASAIVAPA-----AAGCEAASAAAPALMREAGAAA 81
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 541 AAAAAAAAAAGGAGGCGGCGGFGVGLGGFGGFGGSSAAAAA-----AAGCGG 596
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 SAEREAPATAVAGDNGSGLRREPRADDFDDEGEFEDEBAAAAAAAAAAIGYRDNLILT 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 597 GG-RKRGRRGRGGDD-CNGASAVAAAAAAGSADVAIAAAAAAAM-YGCG---A 650
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 DGL-LTNGF 149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 651 DGPDFDNGF 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3

US-08-425-069-4
 ; Sequence 4, Application US/08425069
 ; Patent No. 5728810

; GENERAL INFORMATION:

; APPLICANT: Lewis, Randolph V.

; APPLICANT: Xu, Ming

; APPLICANT: Himmann, Michael B.

; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK

TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 FIELD OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 NUMBER OF SEQUENCES: 69
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch
 STREET: 301 No. 5728810th Washington Street
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22046
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/425,069
 FILING DATE: 19-Apr-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1447-106P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 TELEX:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 595 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-425-069-4

```

1      MEDIUM TYPE:  IBM PC compatible
2      COMPUTER:      IBM PC compatible
3      OPERATING SYSTEM:  PC-DOS/MS-DOS
4      SOFTWARE:      PatentIn Release #1.0, Version #1.25
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER:  US/08/317,844B
7      FILING DATE:      04-OCT-1994
8      CLASSIFICATION:  435
9      ATTORNEY/AGENT INFORMATION:
10     NAME:  Murphy Jr., Gerald M
11     REGISTRATION NUMBER:  28,977
12     REFERENCE/DOCKET NUMBER:  1447-105P
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE:  (703) 241-1300
15     TELEFAX:  (703) 241-2848
16     TELEX:  248345
17     INFORMATION FOR SEQ ID NO:  4:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH:  595 amino acids
20     TYPE:  amino acid
21     TOPOLOGY:  linear
22     MOLECULE TYPE:  protein
23     US-08-317-844B-4

```

```

Db 52 SEEVIA-EAGEPEVTEPAEAASGPSESP-SPPAEELPESHAEPYPAGGAPG 109
QY 58 AAGCEAASAAAPALWREAGAAASEREAPATAVAGDNGSGLRREPRAD--DFDD 115
Db 110 EQARAGSDSRAGVSEDAAGNEGAAREPAALENGDAD-----EPSSDPDEFDD 162
QY 116 EGEE-----DEAAAAAAAIYRDN-----LLTDLGLJNGFH 150
Db 163 VSEELLGDVLRQEQADGIDSVIVDNVPOVGPDLKLNKIFSKGKITDNFY 222
QY 151 SCESDDORT-----SHA-----SSDWTPRPRIGPTFYQOH-----LMIGDP 190
Db 223 ---PEDEKTKGYIFLEXASPAHVAVKNAD-----GKIDKQHTFRVNLFTDFDK 271
QY 191 RTILDL--LPETIPPELDWTQIYINILSEPPKRRKRDINTIDAVKLLQECK-- 246
Db 272 YMTIDEMDIPKOPFKRLGNLRY-----LEBA-----ECRDQ 305
QY 247 -KIIVLTGAGVSVSGIDPFRSRDGIYARLAVFPDLPD-----QAMFDEYFRKDR 299
Db 306 YSVI-----FESGD-----RTSIFMNDVKDPVSIERARWTEYVWSPK 345

```

```

RESULT 6
5273901-7
: Patent No. 5273901
: APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
: SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
: AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
: TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
: SPOROZOITE 21.5 KB ANTIGEN, AC-6B
: NUMBER OF SEQUENCES: 11
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/581,693
: FILING DATE: 12-SEP-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 215,162
: FILING DATE: 05-JUL-1988
: APPLICATION NUMBER: 746,520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627,811
: FILING DATE: 05-JUL-1984
: SEQ ID NO: 7:
: LENGTH: 180
5273901-7

```

```

Query Match 3.1%, Score 119, DB 5; Length 180;
Best Local Similarity 33.6%, Pred. No. 0.004;
Matches 41; Conservative 8; Mismatches 41; Indels 32; Gaps 2;

```

```

QY 10 QAAGSPSAAAMEASOPADEPLRKRPRRDGGLGRSGEPSSAAVAPAAAGCEAASAP 69
Db 31 ERRAAPAAATATAAPAAATA-----ATAATAATAATTPAAAAAP 69
QY 70 AALWREAGAAASAREAPATAVAGDNGSGLRREPRADDFDDDEGEDEAAAAAA 129
Db 70 AAAAPAAAAAATGAAAPAAATAAAGAAKAAAEART-----ETEGAGAGAA 118
QY 130 AA 131
Db 119 EA 120

```

```

RESULT 7
5482709-6
: Patent No. 5482709
: APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
: SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
: AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
: TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
: ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS

```

```

: NUMBER OF SEQUENCES: 10
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/148,432
: FILING DATE: 08-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 581,693
: FILING DATE: 12-SEP-1990
: APPLICATION NUMBER: 215,162
: FILING DATE: 05-JUL-1989
: APPLICATION NUMBER: 746,520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627,811
: FILING DATE: 05-JUL-1984
: SEQ ID NO: 6:
: LENGTH: 180
5482709-6

```

```

Query Match 3.1%, Score 119, DB 5; Length 180;
Best Local Similarity 33.6%, Pred. No. 0.004;
Matches 41; Conservative 8; Mismatches 41; Indels 32; Gaps 2;

```

```

QY 10 QAAGSPSAAAMEASOPADEPLRKRPRRDGGLGRSGEPSSAAVAPAAAGCEAASAP 69
Db 31 ERRAAPAAATATAAPAAATA-----ATAATAATAATTPAAAAAP 69
QY 70 AALWREAGAAASAREAPATAVAGDNGSGLRREPRADDFDDDEGEDEAAAAAA 129
Db 70 AAAAPAAAAAATGAAAPAAATAAAGAAKAAAEART-----ETEGAGAGAA 118
QY 130 AA 131
Db 119 EA 120

```

```

RESULT 8
US-08-651-999A-1
: Sequence 1, Application US/08651999A
: Patent No. 6031088
: GENERAL INFORMATION:
: APPLICANT: Stefan Somlo and Toshio Mochizuki
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
: STREET: 90 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10016
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
: MEDIUM TYPE: DISKETTE
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/651,999A
: FILING DATE: MAY 23, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: ELIZABETH A. BOGOSIAN
: REGISTRATION NUMBER: 39,911
: REFERENCE/DOCKET NUMBER: 96700/395
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5995
: TELEFAX: (212) 286-0854 or 286-0082
: TELEX: TWX 710-581-4766
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 866 amino acids
: TYPE: amino acid
: STRANDEDNESS: single

```

TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal fragment
US-08-651-999A-1

Query Match 3.0%; Score 114; DB 3; Length 866;
Best Local Similarity 23.2%; Pred. No. 0.15;
Matches 70; Conservative 31; Mismatches 107; Indels 94; Gaps 13;

QY 26 QPADELRPRRDGPGICRSPGPSAAVAAPAAACEA--ASAAAPALMR-----EAA 77
DB 11 QPGDARPPAPAPDPG-----RLMGCAAGASIAAGIGLCEQGLEIEMQ 57
QY 78 GAAASERAPATAVAGDNGSGLRREPPAADDFDDEGEDEE----- 122
DB 58 RIQAAARPPAGAAASPSPLSCSQAMSRDNPGEAEAEDEEVEEGEGMVEEMDV 117
QY 123 -----AAAAAAAIIGYRDNLITDGLTNGFHSCESDDDRTSHASSSDMTPRPI 174
DB 118 EHRPGSRRAASSAVSVGARSR---GL--GGYHAGHPGSGRRRRREDQGPCSP-V 169
QY 175 GYTFVQOHLMI-GTDPRT-----ILKDLPETIPPP 205
DB 170 GGGDPLRLRLPEGGPPRMAAERLYRGLRGLMGTRLMESSSTRKRYLKSIVREL--- 226
QY 206 ELDDMTLMOIVINILSEPPRRKRKDINTIEDAVKLLQECKIITVLGAGVSVSGIP-D 264
DB 227 ---TYLLFLVICILTYGTREADNRSEI-FYENLLGLVPRIRQLRVNRNG-----SCSIPOD 277
QY 265 FR 266
DB 278 LR 279

RESULT 9

US-08-405-175A-6
Sequence 6, Application US/08405175A
Patent No. 5885772
GENERAL INFORMATION:
APPLICANT: Aderem, Alan A.
APPLICANT: Chen, Jianmin
APPLICANT: Chang, Sandy
TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,175A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-121A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: predicted primary structure of bovine MARCKS
HYPOTHETICAL: NO
US-08-405-175A-6

Query Match 2.9%; Score 113; DB 2; Length 335;
Best Local Similarity 24.1%; Pred. No. 0.039;
Matches 70; Conservative 29; Mismatches 95; Indels 96; Gaps 13;

QY 9 LQAGS-PSA-----AAAEASQPA-----DEPLRRPRRDGPGICRSPGPSAAV----- 54
DB 58 LQANGSPAPADKEEPPAAAGSGAASPAAAKDEPAAA-----APDAGASPEKEAPEGEA 112
QY 55 -----APAAAGCEAASAAAPALMRGAAGAASAREAPA-----TAVAGDNGSGGLRR 104
DB 113 AERGSPTAAEGEAASASTSSPKADGATPSPNETPKKKRFSFKSFKLSGFSFK 172
QY 105 EPPAADDFDDEGE-----EDEAAAAAAAIIGYRDNLITDGLTNGFHSCESDDD 157
DB 173 NKKEAGEGEAEAGACAGASAEKGDEASGGAANA----- 206
QY 158 DRSHASSSDMTPRPIGYTFVQOHLMTGTDPRTILKDLPETIPPELDMTLMQIYI 217
DB 207 ---GEAGAPGEPTAAPGEAAGGEGAGGDPQEAKEPE---EAAVAPE----- 249
QY 218 NILSEPPRRKRKDINTIEDAVKLLQECKIITVLGAGVSVS-----CGIP 263
DB 250 ----KPPARGAK---AVEEPSKAKEAE-----AGVSAGAGCEAP 286

RESULT 10

US-08-405-175A-5
Sequence 5, Application US/08405175A
Patent No. 5885772
GENERAL INFORMATION:
APPLICANT: Aderem, Alan A.
APPLICANT: Chen, Jianmin
APPLICANT: Chang, Sandy
TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,175A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-121A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 332 amino acids
TYPE: amino acid


```

US-08-458-298-2

Query Match          2.9%; Score 111.5; DB 1; Length 832;
Best Local Similarity 27.9%; Pred. No. 0.24;
Matches 60; Conservative 19; Mismatches 103; Indels 33; Gaps 6;

OY 11 AAGSAAANMEAAQDPDEPLRKRRPRDRGPELGSPSPSAVAAPAAAGCAASAAAPA 70
Db 612 AGAGAGAAAAGAGAGGGAG-----GGRGAGAGAAAAGAGAAAG-----AGACA 654
OY 71 ALMREAAGAAASAESEAPATAVAGGDNDSGLRRBPRAADDDDEGEDEDE---AAAA 127
Db 655 GGGGGGGGAGAGAGAAAAGAGAGGCGRG-----AGAGGAGGGGCGAGAGAGAGAAAA 709
OY 128 AAAAGYRDNLLLTDLTLTNGFHCESDDDDRTSHASSSDMTPREPRIGYTFVQOHLMI 187
Db 710 GAGAGGCGYDKEIAC-----WSRCRYVASTTSRLSSAESAASSRISASASTLVSGYLNT 762
OY 188 TDPRITLKDLPET-IPPELDDMTLMOIVINILS 221
Db 763 AALPSVISDLFAQVAGASSPVIRKSRSLIOLLEIVS 797

RESULT 14
US-07-814-220-2
; Sequence 2, Application US/07814220
; Patent No. 5925540
; GENERAL INFORMATION:
; APPLICANT: Caceci, Thomas
; APPLICANT: Toch, Thomas E.
; APPLICANT: Szumanski, Maria B.W.
; TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
; TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WHITHAM, CURTIS & WHITHAM
; STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
; STREET: Suite 900
; CITY: Reston
; STATE: VA
; COUNTRY: USA
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,220
; FILING DATE: 23-DEC-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/588,437
; FILING DATE: 25-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitham, Michael E.
; REGISTRATION NUMBER: 32,635
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-391-9035
; TELEFAX: 703-391-2510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-814-220-2

```

	Matches	42;	Conservative	7;	Mismatches	45;	Indels	38;	Gaps	2.
OY	2	ADEVALIAIQAGSPSNAAMENASQPADEPLKRRRRDGGGLGRSGPGEASAAPAAACG	61							
Dd	5	SDAAAAAATATAATAAAAAATTAA-----TAAAAAAAATTA	38							
OY	62	EASASAAPALMLREAGAASASAREAPATAVAGDGDSGLRPEAADPFDDDEEED	121							
Dd	39	TAAAAAATAATTAATAAAAAATTAATAAAAAATATAAKAAALTAANANA-----	86							
OY	122	EAAAAAAAAAATIG	133							
Dd	87	AAAAATAAAAARG	98							


```

Db 1 MADEVALALQAAGSPSSAAMEASQPADPLRRKRRDGGGLGHSPEGPSAAVAAPAAAG 60
QY 61 CEASNAAPALMLRMAAASAEAREAPATAVAGDNGSGLRREPRADDFDDEGEE 120
    |||||||
Db 61 CEASNAAPALMLRMAAASAEAREAPATAVAGDNGSGLRREPRADDFDDEGEE 120
QY 121 DEAAAAAIAIGYRDNLITDGLITNGFHSCESDDDRTSHASSDWTTPRIRGPTFFV 180
    |||||||
Db 121 DEAAAAAIAIGYRDNLITDGLITNGFHSCESDDDRTSHASSDWTTPRIRGPTFFV 180
QY 181 OOHLMIGDPTILKDLIPETIPPELDMLQIVITLSEPPRRKKRKDINTIEDAVK 240
    |||||||
Db 181 OOHLMIGDPTILKDLIPETIPPELDMLQIVITLSEPPRRKKRKDINTIEDAVK 240
QY 241 LLOECKKTIIVLTGAGVSVSGIIPDRSRDGIYARLAVDFPDLPPQAMFDEYFRAKDRP 300
    |||||||
Db 241 LLOECKKTIIVLTGAGVSVSGIIPDRSRDGIYARLAVDFPDLPPQAMFDEYFRAKDRP 300
QY 301 FFKFAKEIYPOQFOPSLCHKFTALSDKEGKLLRNTQNIIDTLEQVAGIQRILQCHGSPAT 360
    |||||||
Db 301 FFKFAKEIYPOQFOPSLCHKFTALSDKEGKLLRNTQNIIDTLEQVAGIQRILQCHGSPAT 360
QY 361 ASCLCKKVKCEAVRGDIFNOVPRCPCRADEPLAIMKEIYFGENLEBQFIRAKKY 420
    |||||||
Db 361 ASCLCKKVKCEAVRGDIFNOVPRCPCRADEPLAIMKEIYFGENLEBQFIRAKKY 420
QY 421 DKDEVDLLIVGSSLSKVPVALIPSSIPHEVPQILINREPLPHLHFDELLGDCVITNE 480
    |||||||
Db 421 DKDEVDLLIVGSSLSKVPVALIPSSIPHEVPQILINREPLPHLHFDELLGDCVITNE 480
QY 481 LCHRLGGEYAKLCNPKVPLSEITEKPPRQKELVHLSLPTPLHISDSSSPERTVPOD 540
    |||||||
Db 481 LCHRLGGEYAKLCNPKVPLSEITEKPPRQKELVHLSLPTPLHISDSSSPERTVPOD 540
QY 541 SSVATLVDQATNNNVNNDLEVSSESCVEKPEQVOTSRNVENINENDFRAGVSTADK 600
    |||||||
Db 541 SSVATLVDQATNNNVNNDLEVSSESCVEKPEQVOTSRNVENINENDFRAGVSTADK 600
QY 601 NERTSVATVRCMPNRLAKQISKRLEGNOYLEPFPNRYIFHGAEVYSDSEDDVLSSS 660
    |||||||
Db 601 NERTSVATVRCMPNRLAKQISKRLEGNOYLEPFPNRYIFHGAEVYSDSEDDVLSSS 660
QY 661 CGSNSDGTCSPLSEPLEDESEIEEYFNGLEDTERPECAGSGFGADGDDQEVNVEA 720
    |||||||
Db 661 CGSNSDGTCSPLSEPLEDESEIEEYFNGLEDTERPECAGSGFGADGDDQEVNVEA 720
QY 721 IATROELTDVNPSPDKS 737
    |||||||
Db 721 IATROELTDVNPSPDKS 737

```

```

RESULT 2
US-09-461-580A-26
; Sequence 26, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-461-580A-26

```

```

Query Match 100.0%; Score 3854; DB 18; Length 737;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 737; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEVALALQAAGSPSSAAMEASQPADPLRRKRRDGGGLGHSPEGPSAAVAAPAAAG 60
    |||||||
Db 1 MADEVALALQAAGSPSSAAMEASQPADPLRRKRRDGGGLGHSPEGPSAAVAAPAAAG 60
QY 61 CEASNAAPALMLRMAAASAEAREAPATAVAGDNGSGLRREPRADDFDDEGEE 120
    |||||||
Db 61 CEASNAAPALMLRMAAASAEAREAPATAVAGDNGSGLRREPRADDFDDEGEE 120
QY 121 DEAAAAAIAIGYRDNLITDGLITNGFHSCESDDDRTSHASSDWTTPRIRGPTFFV 180
    |||||||
Db 121 DEAAAAAIAIGYRDNLITDGLITNGFHSCESDDDRTSHASSDWTTPRIRGPTFFV 180
QY 181 OOHLMIGDPTILKDLIPETIPPELDMLQIVITLSEPPRRKKRKDINTIEDAVK 240
    |||||||
Db 181 OOHLMIGDPTILKDLIPETIPPELDMLQIVITLSEPPRRKKRKDINTIEDAVK 240
QY 241 LLOECKKTIIVLTGAGVSVSGIIPDRSRDGIYARLAVDFPDLPPQAMFDEYFRAKDRP 300
    |||||||
Db 241 LLOECKKTIIVLTGAGVSVSGIIPDRSRDGIYARLAVDFPDLPPQAMFDEYFRAKDRP 300
QY 301 FFKFAKEIYPOQFOPSLCHKFTALSDKEGKLLRNTQNIIDTLEQVAGIQRILQCHGSPAT 360
    |||||||
Db 301 FFKFAKEIYPOQFOPSLCHKFTALSDKEGKLLRNTQNIIDTLEQVAGIQRILQCHGSPAT 360
QY 361 ASCLCKKVKCEAVRGDIFNOVPRCPCRADEPLAIMKEIYFGENLEBQFIRAKKY 420
    |||||||
Db 361 ASCLCKKVKCEAVRGDIFNOVPRCPCRADEPLAIMKEIYFGENLEBQFIRAKKY 420
QY 421 DKDEVDLLIVGSSLSKVPVALIPSSIPHEVPQILINREPLPHLHFDELLGDCVITNE 480
    |||||||
Db 421 DKDEVDLLIVGSSLSKVPVALIPSSIPHEVPQILINREPLPHLHFDELLGDCVITNE 480
QY 481 LCHRLGGEYAKLCNPKVPLSEITEKPPRQKELVHLSLPTPLHISDSSSPERTVPOD 540
    |||||||
Db 481 LCHRLGGEYAKLCNPKVPLSEITEKPPRQKELVHLSLPTPLHISDSSSPERTVPOD 540
QY 541 SSVATLVDQATNNNVNNDLEVSSESCVEKPEQVOTSRNVENINENDFRAGVSTADK 600
    |||||||
Db 541 SSVATLVDQATNNNVNNDLEVSSESCVEKPEQVOTSRNVENINENDFRAGVSTADK 600
QY 601 NERTSVATVRCMPNRLAKQISKRLEGNOYLEPFPNRYIFHGAEVYSDSEDDVLSSS 660
    |||||||
Db 601 NERTSVATVRCMPNRLAKQISKRLEGNOYLEPFPNRYIFHGAEVYSDSEDDVLSSS 660
QY 661 CGSNSDGTCSPLSEPLEDESEIEEYFNGLEDTERPECAGSGFGADGDDQEVNVEA 720
    |||||||
Db 661 CGSNSDGTCSPLSEPLEDESEIEEYFNGLEDTERPECAGSGFGADGDDQEVNVEA 720
QY 721 IATROELTDVNPSPDKS 737
    |||||||
Db 721 IATROELTDVNPSPDKS 737

```

```

RESULT 3
US-09-461-580A-12
; Sequence 12, Application US/09461580A
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-Ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461.580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12

```

LENGTH: 327
TYPE: PRF
ORGANISM: Mus musculus
US-09-461-580A-12

Query Match 44.9%; Score 1732; DB 18; Length 327;
Best Local Similarity 100.0%; Pred. No. 2,4e-134;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 VINILSEPPKRRKRKDIINTEDAVKLLQECKKIIVLTGAVSVSCGIPDRSRDGIYARL 275
DB 1 VINILSEPPKRRKRKDIINTEDAVKLLQECKKIIVLTGAVSVSCGIPDRSRDGIYARL 60
QY 276 AVDFPPLPDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNY 335
DB 61 AVDFPPLPDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNY 120
QY 336 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP 395
DB 121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP 180
QY 396 LAIMKEIVFEGENLEPEOFHRAKMYKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQIL 455
DB 181 LAIMKEIVFEGENLEPEOFHRAKMYKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQIL 240
QY 456 INREPRPHAFVYELLGDCDYIINELCHRGSEYAKLCNPKVLSSETTEPRPRQELVH 515
DB 241 INREPRPHAFVYELLGDCDYIINELCHRGSEYAKLCNPKVLSSETTEPRPRQELVH 300
QY 516 LSELPTPLHISEDSSPERTVPQDSS 542
DB 301 LSELPTPLHISEDSSPERTVPQDSS 327

RESULT 4
US-09-461-580A-4
Sequence 4, Application US/09461580A
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard
APPLICANT: Imai, Shin-Ichiro
APPLICANT: Armstrong, Christopher
TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
FILE REFERENCE: 0050.1618-000
CURRENT APPLICATION NUMBER: US/09/461,580A
CURRENT FILING DATE: 1999-12-15
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 245
TYPE: PRF
ORGANISM: Mus musculus
US-09-461-580A-4

Query Match 33.4%; Score 1288; DB 18; Length 245;
Best Local Similarity 100.0%; Pred. No. 6,6e-98;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 VINILSEPPKRRKRKDIINTEDAVKLLQECKKIIVLTGAVSVSCGIPDRSRDGIYARL 275
DB 1 VINILSEPPKRRKRKDIINTEDAVKLLQECKKIIVLTGAVSVSCGIPDRSRDGIYARL 60
QY 276 AVDFPPLPDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNY 335
DB 61 AVDFPPLPDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNY 120
QY 336 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP 395
DB 121 TONIDTLEQVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP 180

QY 396 LAIMKEIVFEGENLEPEOFHRAKMYKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQIL 455
DB 181 LAIMKEIVFEGENLEPEOFHRAKMYKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQIL 240
QY 456 INREP 460
DB 241 INREP 245

RESULT 5
US-09-619-049-576
Sequence 576, Application US/09619049
GENERAL INFORMATION:
APPLICANT: YANDELL, MARK
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSECTICIDAL
FILE REFERENCE: C1000735
CURRENT APPLICATION NUMBER: US/09/619,049
CURRENT FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/171,590
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/171,627
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/175,763
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/175,685
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/186,663
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/187,241
PRIOR FILING DATE: 2000-03-03
NUMBER OF SEQ ID NOS: 1533
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 576
LENGTH: 823
TYPE: PRF
ORGANISM: DROSOPHILA
US-09-619-049-576

Query Match 30.3%; Score 1169; DB 20; Length 823;
Best Local Similarity 38.8%; Pred. No. 3,1e-87;
Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

QY 57 AAGCEAASAAAPALMREA-----GAASAERAPATAVAGODNCGLRREPR----- 107
DB 42 AASTTEAFAEAATATTTETATSELAKANGELIKTLAREDEIGANLEHKTKNFTKS 101
QY 108 AADDFDDDEGEDEDEAAAAAAGYRDNLLITDGLITNGFSCESDDDDRTSHASS- 166
DB 102 MGEDEDEDEEEDDEE-----EDD-----EGITGTSNEDESSNCSSV 145
QY 167 --DWTFRPRIGPYTVQOHLMTGPTIILKDLPE--TTPPELDMTLMQIVINILSE 222
DB 146 EPDWKLR-----WLOREPYTGRVPROVIASIMPHRTGLAGTDSDVIMDLAHLNLE 198
QY 223 PPKRRKRKDIINTEDAVKLLQECKKIIVLTGAVSVSCGIPDRSRDGIYARLAVDPDL 282
DB 199 PKRRNKIASVNTEDDVYSLVKSQKIIVLTGAVSVSCGIPDRSRDGIYARLAVDPDL 258
QY 283 PDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNYTONIDTL 342
DB 259 PDPQAMFDEIFPKDPPEFFKFAKEIYPOGFSLCHKFTALSDKSKLLRNYTONIDTL 318
QY 343 EQVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP----- 391
DB 319 ERVAGIQRILQCHGSFATASCLICKYKVCDAVGRDIFNOVVRCPADAP----- 378
QY 392 --ADEPL-----AIMEPELVFEGENLEPEOFHRAKMYKDEVDLLIVIGSSLKVRPALI 443
DB 379 AVTEBELQVLVENGIMKPDIVTFEGELPDEXHYVMATDKVCDLLIVIGSSLKVRPAHI 438

```

444 PSSIPHEVPOILINREPLPHLHFDVLLGDCVYINELCHRLGGE---YAKLCNPKYKLS 500
439 PSSIPATVPOLLINREDLHLKFDVELLGDSDVIINICHRLSNDNDCWMLCCODESVLT 498
501 ETEKPPRPOKELVHSELPLPHISEDSSPER-----TVPQDSSVIATLVDA 551
499 ESKELMP-PESHNNHLH---HLHHRHCSESESRQSLDTQSIKSNSSADYILISAG 554
552 TNNVNDLEVSSESC-----VE-----EKPO-EVQT 576
555 TCSD-SGEFESTSCGKRSTAAEAALERTKIDILVELNETTALSCDRLGLEGPQTIVES 613
577 SRNV-----ENINVENDFKAVGSSSTADKN---ERTVAETVRKCPNRLAKEQISKRL 628
614 YRHLSIDSSKSGIEQCDNEATPSYVRPSNLVQETKVAPSLPPIPOQRGKRYTAERLQ 673
629 -GNQYLFVPPNRYTFHGAEV---YSDSEDDVLSS-----SCGSN---SDSGTC 670
674 PGTFYSHNNYSYVFGAQVFWNDYSDDDDEEERSHNRHSDLFGVGNHYKDDDEDAC 733
671 Q-----SPSLE-----EPLDEDS 683
734 DLNAVPLSPLLPPLSLAHIVTDIVNGSNEPLPNS 768

RESULT 6
US-60-167-217-10885
; Sequence 10885, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10885
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-10885

Query Match 30.3%; Score 1169; DB 22; Length 823;
Best Local Similarity 38.8%; Pred. No. 3.1e-87;
Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

57 AAACCEAASAAAPALMREAA---GAASAERAPATAVAGDNGSGLRREPR----- 107
42 ASTTEAEAEATATTEPATSELAKANGEIKTKTLAAREEDELIGANLEHKTKNPTKS 101
108 AADFDDEDEGEDEEAASAAAAAIGYRDNLLTDLGLTNFHCESDDDDRTSHASS- 166
102 MGEDEDEDEEEDDEBE-----EDD-----EGSITGTSNDEDESSNCSSSV 145
167 --DWTTPRRIGPYFVOOHLMTGTPRTILKDLPE--TIPPELDMTLMQIYINLISE 222
146 EPDMKLR-----WLOREFYTGVRPROVIAISIMPHATGAGTDSVLDYLAHLNE 198
223 PPKRRKRKDINTIEDAVKLLOECKKIIVLTGAGVSVSGIPDFRSRDIYARLAVDPDL 282
199 PKRNNKLASVNTFEDVLSLVKKSQKIIVLTGAGVSVSGIPDFRSTNGITARLHDPDL 258
283 PDPQAMPDIEFRKDPREFEFKAETIYPGQFQPSLCHKFIASLSEKGLLNRYNTQNTDL 342
259 PDPQAMPDINFKRDPREFYKFAREIYGERQPSCHRFIMLETGKGLLNRYNTQNTDL 318
343 EQVAVGIORILLCHGSFATASCLICKYKVDCAVAGDIFNOVAVPRCP----- 391
319 ERVAGIORVIECHGSFSFASCTCRFCNADALRADIFAGRIYVCPQCPQPKESVDASV 378
```

```

392 --ADEPL-----AImKEIYFEGENLPEQFHRAMKXKDEVDLLIYIGSLKVRVALI 443
379 AVTEELRLQVGENIMKMDIYFEGEGLPDEYHYWATMDKVCDDLIIYIGSLKVRVAHI 438
444 PSSIPHEVPOILINREPLPHLHFDVLLGDCVYINELCHRLGGE---YAKLCNPKYKLS 500
439 PSSIPATVPOLLINREDLHLKFDVELLGDSDVIINICHRLSNDNDCWMLCCODESVLT 498
501 ETEKPPRPOKELVHSELPLPHISEDSSPER-----TVPQDSSVIATLVDA 551
499 ESKELMP-PESHNNHLH---HLHHRHCSESESRQSLDTQSIKSNSSADYILISAG 554
552 TNNVNDLEVSSESC-----VE-----EKPO-EVQT 576
555 TCSD-SGEFESTSCGKRSTAAEAALERTKIDILVELNETTALSCDRLGLEGPQTIVES 613
577 SRNV-----ENINVENDFKAVGSSSTADKN---ERTVAETVRKCPNRLAKEQISKRL 628
614 YRHLSIDSSKSGIEQCDNEATPSYVRPSNLVQETKVAPSLPPIPOQRGKRYTAERLQ 673
629 -GNQYLFVPPNRYTFHGAEV---YSDSEDDVLSS-----SCGSN---SDSGTC 670
674 PGTFYSHNNYSYVFGAQVFWNDYSDDDDEEERSHNRHSDLFGVGNHYKDDDEDAC 733
671 Q-----SPSLE-----EPLDEDS 683
734 DLNAVPLSPLLPPLSLAHIVTDIVNGSNEPLPNS 768

RESULT 7
US-60-171-627-915
; Sequence 915, Application US/60171627
; GENERAL INFORMATION:
; APPLICANT: Yandell, Mark
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL
; FILE REFERENCE: CLO00177
; CURRENT APPLICATION NUMBER: US/60/171,627
; CURRENT FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Drosophila
US-60-171-627-915

Query Match 30.3%; Score 1169; DB 22; Length 823;
Best Local Similarity 38.8%; Pred. No. 3.1e-87;
Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

57 AAACCEAASAAAPALMREAA---GAASAERAPATAVAGDNGSGLRREPR----- 107
42 ASTTEAEAEATATTEPATSELAKANGEIKTKTLAAREEDELIGANLEHKTKNPTKS 101
108 AADFDDEDEGEDEEAASAAAAAIGYRDNLLTDLGLTNFHCESDDDDRTSHASS- 166
102 MGEDEDEDEEEDDEBE-----EDD-----EGSITGTSNDEDESSNCSSSV 145
167 --DWTTPRRIGPYFVOOHLMTGTPRTILKDLPE--TIPPELDMTLMQIYINLISE 222
146 EPDMKLR-----WLOREFYTGVRPROVIAISIMPHATGAGTDSVLDYLAHLNE 198
223 PPKRRKRKDINTIEDAVKLLOECKKIIVLTGAGVSVSGIPDFRSRDIYARLAVDPDL 282
199 PKRNNKLASVNTFEDVLSLVKKSQKIIVLTGAGVSVSGIPDFRSTNGITARLHDPDL 258
283 PDPQAMPDIEFRKDPREFEFKAETIYPGQFQPSLCHKFIASLSEKGLLNRYNTQNTDL 342
259 PDPQAMPDINFKRDPREFYKFAREIYGERQPSCHRFIMLETGKGLLNRYNTQNTDL 318
```



```

QY 343 EQVAGIORTIIOCHGSFATASCLICKYKVCBAVRGDIFFNOVPPRCPCP----- 391
D 319 ERVAGIORVIECHGSFSTASCTKCRFKCNADALADIFAQRIYPCQCPKKEOSVDASV 378
QY 392 --ADEPL-----AIMEPEIVFGENLPEOPFHRAKMYDKDEVDLLIYIGSSLKVRPALI 443
D 379 AVTEELRLQVLVENGIMKPDIVFEGGLPDEYHTVATMDKDCDILLIYIGSSLKVRPAHI 438
QY 444 PSSIPHEVPOILLINREPLPHLHFDVLLGDCDVIINELCHRLGSE---YAKLCNPKYLS 500
D 439 PSSIPATVPOILLINREPLPHLHFDVLLGDCDVIINELCHRLSNDNDCWRLCCDESVLT 498
QY 501 EITKPPRPOKELVHLSLPTPLHISEDSSPER-----TVPODSSVIATLVDAQ 551
D 499 ESKELMP-PHNSNHLH---HLHHRHCSSSESRQSLDTQTSIKSNSADYILGSAG 554
QY 552 TNNVNDLEVSSESC-----VE-----EKPQ-EVQT 576
D 555 TCSD-SGFESTFGCGKSTAAEAALERIKTDILVELNETTALSCBRLGLEGQTTYES 613
QY 577 SRNV-----ENINVENPDFKAVGSTADKN---ERTSYAETVRKCPWRLAKEDISKRL 628
D 614 YRHLSIDSSKSGIEOCNENATPSYVRPSNLVQETKTVAAPSLTPIPOGRKQTAERLQ 673
QY 629 -GNQYLFVPPRNYIFHGAEV-----YSDDEDVLS--SCGSN---SDSGTC 670
D 674 PGTYSHTNNYSYVFGAQVWMDYSDDDDEEBERSINRISDLFGVNGHYKDDDEDAC 733
QY 671 Q-----SPSLE-----EPLEDES 683
D 734 DLNAVPLSPLPLPSLEAHIVTDIVNGSNEPLPNS 768

RESULT 8
US-60-173-464-8827
: Sequence 8827, Application US/60173464
: GENERAL INFORMATION:
: APPLICANT: LI, Peter W.D.
: TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
: FILE REFERENCE: CLO00173
: CURRENT APPLICATION NUMBER: US/60/173,464
: CURRENT FILING DATE: 1999-12-29
: NUMBER OF SEQ ID NOS: 30269
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 8827
: LENGTH: 823
: TYPE: PR
: ORGANISM: Drosophila
US-60-173-464-8827

```

```

Query Match          30.3%; Score 1169; DB 22; Length 823;
Best Local Similarity 38.8%; Pred. No. 3,1e-87;
Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

```

```

QY 57 AAGCEAASAPALMREA-----GAASAREAPATAVAGDNGSGLRREPR----- 107
D 42 ASTTEAEAEATATTEPATSELAKANGEIKTKTLAAREDEIGANLEHKTKNPTKS 101
QY 108 AADPFDDDEGEDEEAAAAAAAAAIGYDNLLLDGLTNFHCESDDORTSHASS- 166
D 102 MGEDEDEDEEEDDEE-----ED-----EBSITGTSDNEDSSSSNCSSSV 145
QY 167 --DWTPRRIGRYTVOOHLIGTDPRTILKDLPE--TIPPELDMMTLQVIYINISE 222
D 146 EPDKLR-----WLOREFYTGVRPROVYASIMPHFATGLAGTDDSVLMDYLAHLNE 198
QY 223 PPKRRKRDINTIEDAVKLOECKRIIVLTGAGVSVSCGIPDFSRDQIVARLAVDPDL 282
D 199 PKRRKLASVMTFDVYISLVKSKIIIVLTGAGVSVSCGIPDFSRDQIVARLAVDPDL 258

```

```

QY 283 PPOAMPDIEYERKDPREFEAKETVPGORPSLCHKETALSDEKGLLBNYTONIDTL 342
D 259 PPOAMPDINIFKDPREFYFAREITPGEFOPCHRFIMLTKKLLNNYTONIDTL 318
QY 343 EQVAGIORTIIOCHGSFATASCLICKYKVCBAVRGDIFFNOVPPRCPCP----- 391
D 319 ERVAGIORVIECHGSFSTASCTKCRFKCNADALADIFAQRIYPCQCPKKEOSVDASV 378
QY 392 --ADEPL-----AIMEPEIVFGENLPEOPFHRAKMYDKDEVDLLIYIGSSLKVRPALI 443
D 379 AVTEELRLQVLVENGIMKPDIVFEGGLPDEYHTVATMDKDCDILLIYIGSSLKVRPAHI 438
QY 444 PSSIPHEVPOILLINREPLPHLHFDVLLGDCDVIINELCHRLGSE---YAKLCNPKYLS 500
D 439 PSSIPATVPOILLINREPLPHLHFDVLLGDCDVIINELCHRLSNDNDCWRLCCDESVLT 498
QY 501 EITKPPRPOKELVHLSLPTPLHISEDSSPER-----TVPODSSVIATLVDAQ 551
D 499 ESKELMP-PHNSNHLH---HLHHRHCSSSESRQSLDTQTSIKSNSADYILGSAG 554
QY 552 TNNVNDLEVSSESC-----VE-----EKPQ-EVQT 576
D 555 TCSD-SGFESTFGCGKSTAAEAALERIKTDILVELNETTALSCBRLGLEGQTTYES 613
QY 577 SRNV-----ENINVENPDFKAVGSTADKN---ERTSYAETVRKCPWRLAKEDISKRL 628
D 614 YRHLSIDSSKSGIEOCNENATPSYVRPSNLVQETKTVAAPSLTPIPOGRKQTAERLQ 673
QY 629 -GNQYLFVPPRNYIFHGAEV-----YSDDEDVLS--SCGSN---SDSGTC 670
D 674 PGTYSHTNNYSYVFGAQVWMDYSDDDDEEBERSINRISDLFGVNGHYKDDDEDAC 733
QY 671 Q-----SPSLE-----EPLEDES 683
D 734 DLNAVPLSPLPLPSLEAHIVTDIVNGSNEPLPNS 768

RESULT 9
US-60-191-637-10901
: Sequence 10901, Application US/60191637
: GENERAL INFORMATION:
: APPLICANT: Venter, J. Craig
: TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
: TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME. AND
: FILE REFERENCE: CLO00392
: CURRENT APPLICATION NUMBER: US/60/191,637
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 42660
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 10901
: LENGTH: 823
: TYPE: PR
: ORGANISM: DROSOPHILA
US-60-191-637-10901

```

```

Query Match          30.3%; Score 1169; DB 22; Length 823;
Best Local Similarity 38.8%; Pred. No. 3,1e-87;
Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

```

```

QY 57 AAGCEAASAPALMREA-----GAASAREAPATAVAGDNGSGLRREPR----- 107
D 42 ASTTEAEAEATATTEPATSELAKANGEIKTKTLAAREDEIGANLEHKTKNPTKS 101
QY 108 AADPFDDDEGEDEEAAAAAAAAAIGYDNLLLDGLTNFHCESDDORTSHASS- 166
D 102 MGEDEDEDEEEDDEE-----ED-----EBSITGTSDNEDSSSSNCSSSV 145
QY 167 --DWTPRRIGRYTVOOHLIGTDPRTILKDLPE--TIPPELDMMTLQVIYINISE 222
D 146 EPDKLR-----WLOREFYTGVRPROVYASIMPHFATGLAGTDDSVLMDYLAHLNE 198

```

```

0Y 223 PPKRKKRIDITIDAVKLLLOEKKIITVLGAGSVSCGJIPDRSRGIIYARLAVPEPD 282
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 199 PKRRKKLASVTPEDDVLISLVKRSOKIITVLGAGSVSCGJIPDRSRGIIYARLAVPEPD 258
0Y 283 PDPQAMFIDIEYFKRDPREFKFAKEIYVGOFOPSLCKHFTALSDKEGKLLRNTYONIDL 342
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 259 PDPQAMFIDINFKRDPREFYKFAREIYVGEPOSPCHRFIKMETGKILLRNTYONIDL 318
0Y 343 EOVAGIQILLOCHSFSATSLCKYKVDCENAVGDIFFNOVVRPCRPCP----- 391
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 319 ERVAGIQIVIECHSFSFASCTCKFRKCMADALADIFADIRIPYCPQOCOPNKQSDVASV 378
0Y 392 --ADEPL-----AIMKPEIYFGENPPEOFRHMKKDKQEVNDLIYGSLLVPRPALI 443
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 379 AVTEBELKQLVENIGMKRDIYFGESEGPDEYHYMATDKVCCDILYIGSSLVPRVANH 438
0Y 444 PSSIPHEVPOILINREPLPHLHFDVLLGDOVYIINELCHRLGE--YAKLCCNPKVLS 500
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 439 PSSIPATVPOILINREQLHNLKFDVLLGDSVYIINQICRLSDNDQCMQLOCDSDSVL 498
0Y 501 EITEKPRPQKVELYHLSLPTPLHIEDSSPER-----TYPODSSVATLIVDOA 551
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 499 ESKELMP-PESHNNHLH--HLHNRHCSESEKOSQOLDTPQOSIKSNSSADYIILSAG 554
0Y 552 TNNVNDLVESSC-----VE-----EKKO-EYQI 576
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 555 TCSDD-SGESSTFSCGKRSTAAEAAIERIKTIDLIVLENNETLALSCDRGLEBPQTVTES 613
0Y 577 SRNV-----ENINVENDPFKAVGSSTADKN--ERTSVAETVVKCWPNRILAKEQIKRLE 628
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 614 YRHLSDSSKSGLEQCDQNEATPSYVAPRSLNYQSTKTVARSLPIRQQRKQRTAERLD 673
0Y 629 -GNQYLFVPPKRYIFHGAEV-----YDSDEBDVLSS-----SCGSN--SDSGTC 670
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 674 PGTEFSHTNNYSYVFGAOVFWMDYDSDDEEERSHNRHSDLFGVNGHNYKDDDEDAC 733
0Y 671 Q-----SPLE-----EPLEDS 683
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 734 DLNAVPLSPLLPSPLEAHIYTDIVNGSENEPLPNSS 768

```

```

RESULT 10
US-60-191-681-8549
: Sequence 8549, Application US/60191681
: GENERAL INFORMATION:
: APPLICANT: I.I. Peter, W.D.
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
: TITLE OF INVENTION: USES THEREOF.
: FILE REFERENCE: C1000390
: CURRENT APPLICATION NUMBER: US/60/191,681
: CURRENT FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 30973
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8549
: LENGTH: 823
: TYPE: PRT
: ORGANISM: DROSOPHILA
: US-60-191-681-8549

```

Query Match	30.3%	Score 1169;	DB 22;	Length 823;
Best Local Similarity	38.8%	Pred. No. 3.1e-87;		
Matches 293; Conservative	100;	Mismatches 206;	Indels 156;	Gaps 25

[illegible][illegible]

```

RESULT 11
US-09-270-767-44940
: Sequence 44940, Application us/09270767
: GENERAL INFORMATION:
:   APPLICANT: Homburger et al.
:   TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
:   FILE REFERENCE: File Reference: 7326-094
:   CURRENT APPLICATION NUMBER: US/09/270.767
:   CURRENT FILING DATE: 1999-03-17
:   NUMBER OF SEQ ID NOS: 62517
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ ID NO 44940
:   LENGTH: 286
:   TYPE: PRT
:   ORGANISM: Drosophila melanogaster
:   FEATURE:
:   OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44940

```

Query Match	19.5%	Score 750:	DB 16:	Length 286:
Best Local Similarity	54.4%	Pred. NO. 2.2e-53:		
Matches 149:	Conservative 27:	Mismatches 50:	Indels 68:	Gaps 3:

QY	256	VSVSGCIDPERSRDGIYARLAVDEPDLPDPAQMEDIEYFRKDPDPFFFAKEIYPGQFOP	315
bb	1	VSVSGCIDPFSSTNGIYARLVAHDEPDLPDPAQMEINFRKDPDPFFFAKEIYPGQFOP	60

117: em_estp16:*

118: em_estp17:*

119: em_estp18:*

120: em_estp19:*

121: em_estp20:*

122: em_estp21:*

123: em_estp22:*

124: em_estp23:*

125: em_estp24:*

126: em_estp25:*

127: em_estp26:*

128: em_estp27:*

129: em_estp28:*

130: em_estp29:*

131: em_estp30:*

132: em_estp31:*

133: em_estp32:*

134: em_estp33:*

135: em_estp34:*

136: em_estp35:*

137: em_estp36:*

138: em_estp37:*

139: em_estp38:*

140: em_estp39:*

141: em_estp40:*

142: em_estp41:*

143: em_estp42:*

144: em_estp43:*

145: em_estp44:*

146: em_estp45:*

147: em_estp46:*

148: em_estp47:*

149: em_estp48:*

150: em_estp49:*

151: em_estp50:*

152: em_estp51:*

153: em_estp52:*

154: em_estp53:*

155: em_estp54:*

156: em_estp55:*

157: em_estp56:*

158: em_estp57:*

159: em_estp58:*

160: em_estp59:*

161: em_estp60:*

162: em_estp61:*

163: em_estp62:*

164: em_estp63:*

165: em_estp64:*

166: em_estp65:*

167: em_estp66:*

168: em_estp67:*

169: em_estp68:*

170: em_estp69:*

171: em_estp70:*

172: em_estp71:*

173: em_estp72:*

174: em_estp73:*

175: em_estp74:*

176: em_estp75:*

177: em_estp76:*

178: em_estp77:*

179: em_estp78:*

180: em_estp79:*

181: em_estp80:*

182: em_estp81:*

183: em_estp82:*

184: em_estp83:*

185: em_estp84:*

186: em_estp85:*

187: em_estp86:*

188: em_estp87:*

189: em_estp88:*

190: gb_gss25:*

191: gb_gss26:*

192: gb_gss27:*

193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	709.2	18.3	766	18	AI322487	AI322487 mh28d08.y
C 2	683.2	17.7	847	106	BE285744	BE285744 601096313
C 3	621	16.1	632	3	AA199012	AA199012 mu16e07.t
4	616.2	15.9	790	29	AU051636	AU051636 AU051636
5	611.4	15.8	615	107	BE336141	BE336141 us85g03.y
6	530	13.7	540	111	BE656273	BE656273 UI-M-BHO-
7	527	13.6	527	19	AI390668	AI390668 mu91g05.y
8	514	13.3	542	107	BE375879	BE375879 60129695
9	490.6	12.7	587	29	AU044338	AU044338 AU044338
C 10	486	12.6	510	11	AA762210	AA762210 vv47h03.t
11	483	12.5	534	138	BE989110	BE989110 UI-M-CGCP
12	478.4	12.4	480	3	AA217123	AA217123 mu91g05.t
C 13	478.4	12.4	518	14	AA990042	AA990042 ua56f06.t
C 14	476.8	12.3	552	111	BE692325	BE692325 uw10b12.x
15	468	12.1	479	7	AA465858	AA465858 ve88f01.s
C 16	466.4	12.1	541	111	BE692345	BE692345 uw10d12.x
17	456	11.8	468	136	BE849999	BE849999 uw10d12.y
C 18	432	11.2	443	12	AA792085	AA792085 vp47b04.t
19	427.4	11.0	441	136	BE849979	BE849979 uw10b12.y
20	427	11.0	427	25	AI849827	AI849827 UI-M-RLI-
21	427	11.0	439	7	AA413680	AA413680 vc58h12.s
22	426.6	11.0	949	137	BE883278	BE883278 601509016
23	410.6	10.6	763	138	BE995620	BE995620 UI-M-CGCP
24	410.2	10.6	535	25	AI848611	AI848611 UI-M-AH1-
C 25	401.8	10.4	468	25	AI853874	AI853874 UI-M-BHO-
C 26	398	10.3	431	12	AA792076	AA792076 vp47b06.t
C 27	394.4	10.2	567	29	AU024598	AU024598 AU024598
C 28	394	10.2	394	11	AA756446	AA756446 vv53c06.t
C 29	389.8	10.1	395	20	AI465526	AI465526 mu91g05.x
30	380	9.8	392	9	AA549596	AA549596 vk76e11.s
C 31	380	9.8	402	10	AA673258	AA673258 vp49d05.t
C 32	375	9.7	421	92	AA548525	AA548525 L0038A06-
C 33	369.2	9.5	381	147	W62273	W62273 m097b03.t1
34	362	9.4	484	97	AW996552	AW996552 OV3-BM004
35	359.6	9.3	559	93	AM659060	AM659060 145979 MA
36	358.4	9.3	519	21	AI548934	AI548934 UI-R-C3-E
C 37	350	9.0	519	18	AI324125	AI324125 mh28d08.x
C 38	348.2	8.9	392	138	BF012144	BF012144 us38g11.y
C 39	346.2	8.9	414	91	AM504399	AM504399 UI-HF-BNO
40	346.2	8.9	377	7	AA461259	AA461259 zx63a06.t
41	337	8.7	360	9	AA549597	AA549597 vk76e12.s
42	333.8	8.6	392	24	AI751814	AI751814 cl12a08.y
43	332.8	8.6	412	29	AU043723	AU043723 AU043723
C 44	313.6	8.1	332	133	BB544226	BB544226 BB544226
C 45	310.2	8.0				

ALIGNMENTS

RESULT 1
AI322487/c AI322487 766 bp mRNA EST 23-DEC-1998
LOCUS mh28d08.y1 Soares mouse placenta, 4NDMP13.5 14.5 Mus musculus CNA
DEFINITION clone IMAGE:443823 5' similar to WP:R11A8.4 CE06302 YEAST
REGULATORY PROTEIN SIR2 LIKE ; , mRNA sequence.
ACCESSION AI322487
VERSION AI322487.1 GI:4056916
KEYWORDS EST.

SOURCE	mouse mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 766)
AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelasing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. MG1:269135 This read is a RESEQUENCE of a previously sequenced mouse clone correct orientation) Possible reversed clone: similarity on wrong strand Seq primer: -40RP from Gibco High quality sequence stop: 406. Location/Qualifiers 1..766 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone IMAGE:443823" /clone_lib="Soares mouse placenta 4NDMP13.5 14.5" /sex="unknown" /tissue_type="placenta" /dev_stage="adult" /lab_host="DH10B" /note="Organ: Placenta; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI, 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', TGTTACCAATCTGAAGTGGGCGGCGCGGAATTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	220 a 158 c 167 g 217 t 4 others
ORIGIN	
Query Match	18.3%; Score 709.2; DB 18; Length 766;
Best Local Similarity	96.9%; Pred. No. 2.4e-146;
Matches 742; Conservative 0; Mismatches 22; Indels 2; Gaps 2;	
Oy	809 ttggagttctgtctcccttgtygattcctcgactcaagtacaagaagcgtatcatgctcg 868
Db	764 TGGGGTTTCCTGTCTCTCTGTGNAATTCGTGACACAAGAATCAAGGCCGGATTCTTAGTCTG 705
Oy	869 ccttcgggtgacctcccgaccctccagaacctcaagcactgatgttataggatatt 928
Db	704 CC-TCCGGTGCAC-TCCAGACCCTCCACAGCCCTCCAGCCATGCTCATATTGAGATTAT 647
Oy	929 tagaaagacccaagaacctcttcaaglttgcaaaggaatatatacccgcgaagtcca 988
Db	646 TAGAAAGACCCAAGAACCTTCTCTCAAGTTTCCAAGAAAATAATATATCCGACAGTCCA 587
Oy	989 gccgcctctgtgtcaccaatcatcatgcttgcagataagaagaagaaaactacttcgaaa 1048
Db	586 GCCGCTCTGTGTCCAAATTCATAGCTTTGTCAAGATAAGGAAGAAAACTACTCTCGAAA 527

FEATURES	source
ORGANISM	Mus musculus
ORGANISM	house mouse.
ORGANISM	Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM	1 (bases 1 to 847)
ORGANISM	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .
ORGANISM	National Institutes of Health, Mammalian Gene Collection (MGC)
ORGANISM	Unpublished (1999)
ORGANISM	Contact: Robert Strausberg, Ph.D.
ORGANISM	Tel.: (301) 496-1550
ORGANISM	Email: Robert.Strausberg@nih.gov
ORGANISM	Tissue Procurement: Lohar Henthhausen Ph.D., Robin Humphreys
ORGANISM	cDNA Library Preparation: Life Technologies, Inc.
ORGANISM	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
ORGANISM	DNA Sequencing by: Incyte Genomics, Inc.
ORGANISM	Clone distribution: MGC clone distribution information can be
ORGANISM	found through the I.M.A.G.E. Consortium/LLNL at:
ORGANISM	http://image.llnl.gov
ORGANISM	Plate: L1AM8534 row: 1 column: 03
ORGANISM	High quality sequence stop: 733.
ORGANISM	Location/Qualifiers
ORGANISM	1. 847
ORGANISM	/organism="Mus musculus"
ORGANISM	/strain="C57/B6"
ORGANISM	/db_xref="taxon:10090"
ORGANISM	/clone="IMAGE:3491018"
ORGANISM	/clone_11b="NCI_CGAP_Mam5"
ORGANISM	/tissue_type="tumor, gross tissue"

```

/dev_stage="7 months"
/lab_host="DHI08"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by life technologies. Investigators
providing samples: Lotmar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT      225 a      192 c      178 g      252 t
ORIGIN

Query Match      17.7% Score 683.2; DB 106; Length 847;
Best Local Similarity 96.4% Pred. No. 1.4e-140;
Matches 752; Conservative 0; Mismatches 23; Indels 5; Gaps 5;

Oy 729 aaaaagaaagatcaatcaatgaatgctgtaagttactgcagaggtgtaaaag 788
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 780 AAAAGGAAAGATATCCAAACAGAAAGAGGCTGTGAAGTTACTGCAAGAGAGTAAAAAG 721

Oy 789 ataagttctgactggaagctgggggtctgctcct-gtggattcctgactcaagat- 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 720 ATATAGTCTTGACGAGAGCTGGGGTTTCACTCTCTGGGGGATTCCTGACTTCAGATC 661

Oy 847 caagaagcagatctatgctcg-cttgcggtggac-ttccgaacctcccaagacctca 904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 660 CGAGAGAGCGATATGCTATGCTGCGCTTGGCGGTGAGATTCCCAAGCTCCAGACCTCA 601

Oy 905 agccatgttataatgataatgataatgataatgataatgataatgataatgataatg 963
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 AGCCATGTTTCAATATGATGATTTTAAAGAAAGACCCCAAGCCATTCTTCAAGTTTGCA 541

Oy 964 aggaataataatcccgagcagttccagcgcgtctgtgtacaaatcataagcttgtag 1023
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 540 AGGAAATATATCCCGGACAGTTCAGCGCTCTGTGTCAAAATATCATATGCTTGTGAC 481

Oy 1024 ataaggaagaaactactcttgaataataataataataataataataataataata 1083
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 ATAAAGGAAGGAAACTACTCTTGAATAATATATCAAAATATAGATACCTTGAGACAGG 421

Oy 1084 cagaataccaaagagatccctcagtgctcgtctcttgcacagacatctgctgattt 1143
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 CAGGAATCCAAAGATCTTCAAGTGTGATGCTCTTTCGCAACAGCATCTTGCTGATTT 361

Oy 1144 gtaatacaaaagtgtatgtgaagcgtctgctgtagagacatttcaatcaggtagcttc 1203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 GTAATATCAAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 301

Oy 1204 ggtgacctagtgccagcgtgtagagcactgacatcagtagaagccagagattgtctt 1263
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 GGTGCTTGGTGGCCAGCTGATGAGCCACTTGCATCATGAAACCCAGAGATTGTCTTCT 241

Oy 1264 ttggtgaaaacttcaacgaacagtttcatagagcagtagaagtcagcaaaagatgag 1323
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 TTGGTGAAGAACTTACCAAGAAAGTTTCATAGAGCCATGATATGACAAAGATGAAGTT 181

Oy 1324 acctcctcatgttatgtgagatctctctcgtgaagtgtagaccagtagtaactaa 1383
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 ACCCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121

Oy 1384 ctatacccatgtagtgctcctaataataataataataataataataataataataata 1443
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 CTATACCCCAAGAGAGCTCTCAATATATTAATTAATAGGGAAGCTTTGCTCATATTA 61

Oy 1444 ttgagtcagagctcctgtagagcgcagatgtataatgaagttgtcatagagtag 1503
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 TTGATGTAGAGCTCTCTGGAAGACTGCCATGTATTAATTAATGAAGTTGTGTATAGGCT 1

RESULT 3
AA199012/c 632 bp mRNA EST 19-FEB-1997
LOCUS mul607.r1 Soares_thymus_2NDMT Mus musculus CDNA clone IMAGE:639588
DEFINITION 5 similar to WP:R11A8.4 CE06302 YEAST REGULATORY PROTEIN SIR2 LIKE

```

```

ACCESSION      ; mRNA sequence.
AA199012
VERSION        AA199012.1 GI:1794507
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 632)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:391580
Seg primer: -28M13 rev2 from Amer sham
High quality sequence stop: 418.
Location/Qualifiers
1. 632
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:639588"
/clone_1id="Soares_thymus_2NDMT"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DHI08"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
TGTACCAATCTGAGAGGAGCGCGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
3'; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bernaldo."
BASE COUNT      184 a      129 c      137 g      182 t
ORIGIN

Query Match      16.1% Score 621; DB 3; Length 632;
Best Local Similarity 99.8% Pred. No. 7.2e-127;
Matches 632; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 828 ggaattcctgaactcaataaagaagagatcatcgtcgtcggtggagttccca 887
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 632 GGGATTCCTGACTTCATCAATCAAGACGATATCTATGCTCTC-CTTGGGCTGACTTCCCA 574

Oy 888 gacctccagaacctcaagcagatgttgatattgaattttaaagaagaccagaacca 947
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 GACTTCCCAAGCCTTCAAGCATTTGATATTTAGATTTTAAAGAACCCAGACCA 514

Oy 948 ttcttcaagtttgcaagaataatatacccgagacagttccagccgctctgtgtacaaa 1007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 513 TTCTTCAAGTTTGCAAGGAATATATCCGAGACAGTTCCAGCGCTCTGTGTGCACAAA 454

Oy 1008 ttctaagcttgtcagtaagaagaagaactactctgaaatatataactcaaatatgat 1067
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 453 TTCTATGCTTTGTGATGATGAAGGAAGAACTACTTGGAAATATATCTCAAAATATGAT 394

```


[illegible]

Query Match	15.98;	Score 616.2;	DB 29;	Length 790;
Best Local Similarity	91.48;	Pred. No. 8.7e-126;		
Matches 727; Conservative	0;	Mismatches 57;	Indels 11;	Gaps 8

OY	1697	agcaacaacaaatggttaagtattagaagatcctgaatccaagtctgttgaaagaaa	1756
Dn	1	AGCAACAACAAACAACTTTATGATTGTAGAATCATGAAGTGTGTGGAAAGAAA	60
OY	1757	accacaagaatacacactagtagtgaagatctgagaacattaatctggaataatccagatt	1816
Dn	61	ACCACAAGAAGTACAGACTAGTAGAATGTGTGAACATTAAATGTGGAAAAATCCAGATT	120
OY	1817	taagtcgtttgtgttcagctactgcagacaaaaatgaagaacctcagttgcagaacagtt	1876
Dn	121	TAAAGCTGTTGTTGCCAGTACTCGACAACAAAAAATGAAMACOTTCAATTGCAGAAACAGT	180
OY	1877	gagaaatatgtctgcctaataactctcagaaggagacgatattaagcggttgaggagtaa	1936
Dn	181	GAGAAATAGCTGGCTTAATGACTTCGAAAGGACGAGATTAGTAAGCGGCTTGAGGCTPA	240
OY	1937	tcaataaccttctgtaccaccaaatcgtttacataatcccacggctgcctgagataactcaga	1996
Dn	241	TCAATACCTGTTGTACCAACAATAAGTTACATATTCACAGGTCGTGAGGTATACTCAGA	300
OY	1997	ctctggaagatgaacgtcttgtccctcagtctccgtgtggcagtaaacagtgtgacagt	2056
Dn	301	CTCTGGAAGATGACGCTCTTGCCCTCTGATTCCTGTGGCAATCAAGTGCACAGTGGCACATG	360
OY	2057	ccagagtcceaagtttagaanaaccccttggagaagatgaagaatttgaagaattitaca	2116
Dn	361	CCAAGTCCCAAGTTTAGAAGAACCCCTTGGACAGTGAAGATGAATTTGAAGAAATTTACAA	420
OY	2117	tgacttggaagatgatcaggaagagcccgcaatgtgtctgtagagatcgtgaatttgaagctga	2176
Dn	421	TGCGTTGGAAGATGATATCGAGAGAGGCCCGAATGTGCTGGAGAGATCGATTGGAGCTGA	480
OY	2177	tggaggggagatccaagaagttgtttaatgaagctatagctacaagaacaggaatttgaacagatgt	2236
Dn	481	TGGAGGGGAGATCAAGANNTTGTATCAAGCTTN-GCTCAACAAACAGGAATTACAGAGATG	539
OY	2237	aaactatcatcctagacaacatcataaacacattgaagctgtccggatttcaggaattgctgcc	2296
Dn	540	AAACTATACC-TCAGACAAATCATTAACCTCTATTGAA--CTGCCGGAATTCAGAAATGCT-C	595
OY	2297	accagcatctgggaactctgaatcgtlcaaaaaaaatgaatgtttac--tttgaacttgac	2354
Dn	596	ACCAAGCATGTGGAACTTATNCATTGTGCAAAAAATGAATGNTTACCTTGNNACCTTTGACC	655
OY	2355	aaggaaatctgaaagatgtatt--attatagaatcygaaatatagatgtctctctgata	2412
Dn	656	AAGGAATATCGAANAANGWTTTATTATTATTTAGACTCGGAAATNMAATGGCNTTNTNGGGTA	715
OY	2413	attcttaaatctatcatattctgtgttgaactctgtatcatcttcaaacactgtgttgaactc	2472
Dn	716	AATTTCTAAGANNCCATTAATTCNNGTGGACCTTG-NCCTTTAACACTGANNGTGGAC-TC	773
OY	2473	atcttccttccaagy 2487	
Dn	774	ATNTTCTCTTNAGG 788	
RESULT	5		
LOCUS	BE336141	615 bp	mRNA EST 14-JUL-2000

DEFINITION	u8559g3.y1 Soares mouse 3bDMS Mus musculus cDNA clone IMAGE:33252045', mRNA sequence.
ACCESSION	BE336141
VERSION	BE336141.1
KEYWORDS	GI:9209917
SOURCE	EST.
ORGANISM	house mouse. Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurgonathi; Muridae; Murinae; Mus. 1 (bases 1 to 615)
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CCAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov This clone is available royalty-free through LNC ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:1069368 Seq primer: -40RP from Glibco High quality sequence stop: 453. Location/Qualifiers 1..615
FEATURES	
source	

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3325204"
/clone_lib="Soares mouse 3bDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notice="Vector: pPT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15.
TGTATCCAAATCTGTAAGTGGAGCGCGCGCTGTATTTTTTTTTTTTTTTT
3'; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7n3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

```

Query Match	15.8%	Score 611.4	DB 107	Length 615
Best Local Similarity	99.5%	Pred. No. 9.5e-125		
Matches 612	Conservative	0	Mismatches 3	Indels 0
				Gaps 0
QY 3035	cacaatactgccaaagatcgtgaatcgaagaagcccttcgaatcctaataatgctactct	3094		
Db 1	CACAAATACGTGCAAGATGTCATATGCAAAAGCCCTTTGTAATCTTAATTAATGTCATCTT	60		
QY 3095	actgggagagtgtaataattcggactcgtctttccatlaataggaagaacaaatagc	3154		
Db 61	ACTGGGAGAGTGTAAATTTTGGACTGCTGTTTTCCATTAATAGGAAGAACATAGGC	120		
QY 3155	ctcttaattaaagtcaccaagtcataagaataatgtgactccaaccagaagaatcacgtg	3214		
Db 121	CTCTTAATTTAAAGTCCCAAGTCATTAAGATTAATGTAGTCACCAACGAGAAGTACACGT	180		
QY 3215	tgccctgttgaggaattcgtgtaatgtaatcccaagtgcttagcctgtgtaattabgagatga	3274		
Db 181	TGCCCTGTGAGAGATTGGGTGAATGATATATCCCAAGGTGTACGCTGTATTAGGAGATGA	240		
QY 3275	atacagatccaaatagtcataatgaaactcgtctcgtatatttaaagcttagctgcctt	3334		
Db 241	ATACAGATCCAAATAGTCAAAATGAAGACTGTCTTGTATTATTAAGAGCTTAGCTTCCTT	300		
QY 3335	aaaactaggaatcaattcttccaactgagaagaacttttagccttccaacagttcaaac	3394		

Db	301	AAAACGTGGGATCAATTTTCGCAACGTGCAGAAACCTTTGACCTTTCAACAGTTCACAC	360
Qy	3335	tcaaaagtcagttatttattttacagactcttttgaaacattggccccaatttaatat	3455
Db	361	TCACAAAGTCAGTATTTATTTTAAACAGCTCTTTGGAAACCTTGCACCAATTTAAAT	420
Qy	3455	tcacgtgggtttagtatttatacaaaaaaatgatttgaataatagctgtctcttatgca	3514
Db	421	TCATGTGGGTTTACTATTTTATTTCAAAAAAATGATTTGAAATATACGTCTTCATTATGCA	480
Qy	3515	taaaatcccaagttagagacattactgcagagagaaaaagtaattagtagtcaattcc	3575
Db	481	TAAATATCCCGATTAGGACCTTTACTGCGCAGAGAGAAAGTTATAGTAGCTCATTTCC	540
Qy	3575	ctaccctaaagaataactgaaatttatttgcgtacactaaagaatgcatlataatgtttt	3634
Db	541	CTACCTTAAAGATTAACGAATNTATTGGCTACACTTAAAGAAATGACAGTATATTTAGTTTT	600
Qy	3635	ccatttgcatgattg 3649	
Db	601	CCATTTCGATGATGT 615	

RESULT	6
BE656273	
LOCUS	BE656273 540 bp mRNA EST 06-SEP-2000
DEFINITION	UI-M-BH0-ajf-g-10-0-U1.r1 NIH BMAP M_S1 Mus musculus cDNA clone.
ACCESSION	UI-M-BH0-ajf-g-10-0-U1.5', mRNA sequence.
VERSION	BE656273
KEYWORDS	BE656273.1 GI:982186
SOURCE	EST.
ORGANISM	house mouse. Mus musculus.

REFERENCE	1 (Pages 1 to 540)
AUTHORS	Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Chin, H

National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
cDNA Library Preparation: M. B. Soares lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.

```

FEATURES
SOURCE
Location/Qualifiers
1. .540
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="U1-M-BH0-ajf-g-10-0-U1"
/clone_id="NIH_BMAP_M_S1"
/dev_stage="27-32 days"
/lab_host="Dh10B (life technologies)"
/note="Vector: pT7RD-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH BMAP M_S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The driver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of

```


RESULT 8
BE375879 542 bp mRNA EST 21-JUL-2000
LOCUS 601229695F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3593687 5',
DEFINITION mRNA sequence.
ACCESSION BE375879
VERSION BE375879.1 GI:9321244
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 542)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM8766 row: n column: 24
High quality sequence stop: 534.
Location/Qualifiers
1..542
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3593687"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; Salt:
Site:2; Notti: Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 178 a 92 c 140 g 132 t
ORIGIN

Query Match 13.3%; Score 514; DB 107; Length 542;
Best Local Similarity 99.6%; Pred. No. 3.2e-103;
Matches 536; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1762 aagaagtagacacagtaggaatgttggaacatgaatgtgaaatccagatttaagg 1821
|||
DB 5 AAGAAGTACACACTAGTGAATGTGGAACATTAAATGCGAATAATCCATTTTAAGG 64
|||
QY 1822 ctgttggtccagtagctacagacaaaatgaagaacttca-gttgcagaaacagtga 1880
|||
DB 65 CTGTTGGTTCAGTACTGACAGCAAAAATGAAGAATCTTCAAGTTGAGGAGTAAAG 124
|||
QY 1881 aatgctggcctaatagacttgcagaaagagcaga-ttagtaagcggcttgaggtatca 1939
|||
DB 125 AAATGCTGGCTAATAGACTTTGCAAAAGAGCAGATTAGCGGCTTGAGGGTAAATCA 184
|||
QY 1940 ataccgtttgtacacacaaatcgttacatattcaacgtgtcgtgagtatactagactc 1999
|||
DB 185 ATACTGTTTACCACCAATTCGTTACATATTCCACGCTCTGAGGATATCTCAGACTC 244
|||
QY 2000 tgaagatgaagctctgttcctctagttccgtgagcagtaacagtgacagtgacatgca 2059
|||
DB 245 TGAAGATGACGCTTTGTCCTCTAGTCTGTGGCAGTAAAGTGAAGTGGACATGCA 304
|||
QY 2060 gagtcacagtttagaagaaccttggaagatgaagtgaaatgaagaatctcaatgg 2119
|||

DB 305 GAGTCCAAAGTTTAGAAGAACCCCTGGAAAGTGAATGAATGCAATTCACAAATCG 364
QY 2120 ctggaagatgatacaggaagagcccaatgtgtcgtgagagatcgtgattgagcgtgag 2179
|||
DB 365 CTTGGAAGATGATACGAGAGAGCCCGCAATGTGCTGAGAGATCTGGATTGGAGCTGATGG 424
|||
QY 2180 agggatcaagaaggtgtttaaagctatagctacaagaagaatgacagatgttaa 2239
|||
DB 425 AGGGATCAAGAGGTTGTAATGAAGCTATAGCTCAAGACAGGAATGACAGATGTAA 484
|||
QY 2240 ctatccatcagacaaatataacactatggaagctgtccgattgagaattgtccca 2297
|||
DB 485 CTATCCATCAGACCAATATTAACACTATTGAAGCTGTCCGATTTCAGGAATTGCTCCA 542
|||

RESULT 9
A0044338/c 587 bp mRNA EST 09-DEC-1998
LOCUS A0044338 Mouse sixteen-cell-embryo cDNA Mus musculus cDNA clone
DEFINITION J0918D05 3', mRNA sequence.
ACCESSION A0044338
VERSION A0044338.1 GI:3979892
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 587)
AUTHORS Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., Depalma, G.E.,
Liang, Y., Karqul, G.T., Sharara, R., Lim, M.K. and Doi, H.
Systematic analyses of genes expressed in 16-cell mouse embryo (The
ERATO/Doi Project at Wayne State University)
JOURNAL Unpublished (1998)
COMMENT Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBC Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@bioa.jst.go.jp.
Location/Qualifiers
1..587
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0918D05"
/clone_lib="Mouse sixteen-cell-embryo cDNA"
/dev_stage="sixteen-cell-embryo"

BASE COUNT 224 a 85 c 90 g 186 t 2 others
ORIGIN

Query Match 12.7%; Score 490.6; DB 29; Length 587;
Best Local Similarity 97.2%; Pred. No. 4.8e-98;
Matches 530; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 3293 aatgaactagttcttagtatttaaaagcttagcttgccttaaaactagag-atcaatt 3351
|||
DB 582 AAGAAGTACTGCTTGTGATTAATAAGCCTTAGCTGCGCTTAATAACTGGGAATCAATT 523
|||
QY 3352 ttccaactgcagaacttttagcccttcaaaagttcaacacctcagaagaatcagattt 3411
|||
DB 522 TTCTCAACNGCAGAAACTTTTACCTTTCAAAAGTTCCACACCTC--GAAGTCAGTATT 465
|||
QY 3412 atttaccagactctcttggaacattgcccccaaatttaaatatcatcattggttagtat 3471
|||
DB 464 ATTTTACGACTTTCTTGGGACATTTGCCCAATTTAAATTAATCAATGAGGTTTAGAT 405
|||
QY 3472 ttattacaaaaaaatgattgaatatagctgttcttlatgcatataaataaccagtttag 3531
|||
DB 404 TTATTACAAAAAA--GATTGAAATATAGCTGCTTTATGCAATAAATACCGATTAGG 346
|||
QY 3532 accattactgcagagagaaagatlaagtagtcaattccctactactaaagaataact 3591
|||
DB 345 ACCATTACTGCCAGAGGAGAAAGTATTAAAGTACTCATTTCCCTACCTAAAGATACT 286
|||

OY	3592	gaattatcttgccacacaaagaatgcagatataggittcattcgactgcatgtt	3651
Db	285	GAAATTATTTGGCTTCACTAAGAAATGCAGTATATTGTTTTCCATTTCATGATGTGT	226
OY	3562	tctgtctaagacaatacttaaatggaaaaatttgtttaaatattattttaccagtgaag	3711
Db	225	TGTGCTATAGACAAATATTTTAAATTTGAATAATTGTTTTAAATATATTTTACAGTAGA	166
OY	3712	actgttttcagctcttttttatattgtcaatagactttatgtaacctggcatagttttg	3771
Db	165	ACTGTTTTCAGCGCTTTTATATATTTGATACATACACTTTATATGTAATCTGCATATGTTTGG	106
OY	3772	tagaccgtttaatgactgattatcttcctccaacttttgaaaatacaaacagtgcttt	3831
Db	105	TGACCCTTTAATGACTGATATATCTTCNCACAATTITGGAATATCAAACAAGTGTTTT	46
OY	3832	atact 3836	
Db	45	ATACT 41	
RESULT	10		
LOCUS	AA762210/c		
DEFINITION	AA762210	510 bp mRNA EST 27-JAN-1998	
VERSION	vva7fh03.i21 Soares_thymus_2NBMT Mus musculus cDNA clone		
KEYWORDS	IMAGE:1225589.5' similar to WP:RIIAB.4 CE06302 YEAST REGULATORY PROTEIN STR2 LINE ; , mRNA sequence.		
ACCESSION	AA762210		
VERSION	AA762210.1 GI:2811957		
SOURCE	EST.		
ORGANISM	house mouse. Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	I (bases 1 to 510) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubnue,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelning,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The Mashu-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:651181 Possible reversed clone: similarity on wrong strand Seq primer: -28mJ revZ ET from Amersham High quality sequence stop: 63. Location/Qualifiers 1..510 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_image="IMAGE:1225589" /clone_id="Soares_thymus_2NBMT" /sex="male" /issue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B" /note=Vector: pVT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCGCTTTTGTGTGTGTGTGTGTGTGTGTGT 3'] double stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I		

[illegible]

Email: MEST@mail.nih.gov
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
 clones from RESEARCH GENETICS. It should be noted that Bento Soares
 is generating a small number of additional specialized
 non-redundant arrays of BMAP cDNAs whose availability will be
 considered under appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA-No.

FEATURES

source location/Qualifiers
 1..534
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-CG0P-bnk-a-08-0-UI"
 /clone_lib="NIF-BMAP_Ret4_S2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIF-BMAP_Ret4_S2 library is a subtracted library,
 ultimately derived from mouse retina tissue libraries at
 various stages of development. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu.
 TAG_SEQ=None found"
 BASE COUNT 156 a 116 c 117 g 143 t 2 others
 ORIGIN

Query Match 12.5%; Score 483; DB 138; Length 534;
 Best Local Similarity 99.8%; Pred. No. 2.2e-96;
 Matches 483; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

557 gggggcgggagtagtccatattcttggtcagacaatctcagatggaccgaccc 616
 1 GGGGGCGGGAGTAGTCCATATTCTTTGTCAGCAACATTCATGATGGCCGATCC 60
 617 tcgaacaattcttaagaatttaccagaacaattccctcagctgagctgagatc 676
 61 TCGAACAATTCTTAAAGATTATTCAGAAACATTCCTCCAGCTGAGCTGATGAT 120
 677 gaagctgtgcagatgttattatataatccctcagaaccacccaagcgagaaaaa 736
 121 GACGCTGTGGCAGATTGTATTAATATCTTTCAGAACACCAAGCGAAAAA 180
 737 agatacaacaattgaagaatgctgtgaagtctgagagagtgtaaaagatatagt 796
 181 AGATTAACAATGAAGAGAGCTGTGAAGTTACTGCAAGAGTGTAAAGATATAGT 240
 797 tctgactgagcgtgggttctgtctcctgtggaattcctgactcagatcaagaagc 856
 241 TCTGACTGAGCTGGGGTTTCTGTCTCTGTGGGATTCCTGACTCAATCAAGAGCG 300
 857 tatctatgctcgtcctgtcggtgagctccagacccccaagccctcaagcattgt 916
 301 TATCATCTGCTGCTTGGGTGGAGTCTCCAGACCTCCAGACCTCAAGCATTTTGA 360
 917 tatgaagattttgaaagaaccagacattctcaagtttgaaagaataatattcc 976
 361 TATTGAGATTATTTAGAAAGACCCCAAGACCATTTCTCAAGTTTGCAAGGAATATATCC 420
 977 cggagagcttcagacgctcctgtgtcacaaattcagactgtgtcagaataagga 1036
 421 CGGACAGTTCAGCGCGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
 1037 acta 1040
 481 ACTA 484

RESULT 12
 AA217123

LOCUS AA217123 480 bp mRNA EST 06-FEB-1997
 DEFINITION mu91g05.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
 IMAGE:652952 5', mRNA sequence.
 ACCESSION AA217123
 VERSION AA217123.1 GI:1826088
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 480)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:398800

Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 457.
 Location/Qualifiers
 1..480

FEATURES

source location/Qualifiers
 1..480
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:652952"
 /clone_lib="Soares mouse lymph node NBMLN"
 /sex="male"
 /tissue_type="lymph node"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 15',
 TGTACCAATCTGAAGTGGAGCGCGGCGATCTTTTCTTTTCTTTTCTTTTCTTTT
 3'; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73D vector. RNA
 provided by Dr. Bertrand Jordan. Library constructed and
 normalized by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 156 a 77 c 81 g 166 t
 ORIGIN

Query Match 12.4%; Score 478.4; DB 3; Length 480;
 Best Local Similarity 99.8%; Pred. No. 2.3e-95;
 Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3199 ccagaagatactgttccgttgaagattgtgtatgtatcccaagtgtagcct 3258
 1 CCAGAAAGTACACTGTTCCGTGTGAGATTTGGTATGATATCCCAAGGTATGCTT 60
 3259 tgtattatgagatgaataacagatccatagtcataaagaactagttctgtatcaa 3318
 61 TGTATTATGAGATGAATACAGATCAATGATCAATGAATGAATGATTTTAA 120
 3319 aagcttagcttgccttaaacataggatcaatttctcaactgcagaactttagcct 3378
 121 AAGCTTAGCTTGCCTTAACAAGGATCAATTTTCTCAACTGCAGAACTTTAGCCTT 180
 3379 tcaaacagttcacacctcagaagaagtcagattattattacagactccttggaaatgc 3438

Db 181 TCAACAGTTCACACCTCAGAAAGTCAGTATTTATTTACAGACTCTTTGAAACATTCG 240
 QY 3439 ccccaatttaataatctctggtgtttagtattttacaaaaaatgattgaaat 3498
 Db 241 CCCCAATTTAAATATTCATGTGGTTTATTTATTCAAAAAATGATTTGAATAT 300
 QY 3499 agctgtcttataatgaataaccagcttagagcattactccagagagaagaat 3558
 Db 301 AGCTGTTCTTTATGCAATTAATCCAGTTAGACCATTAATCTCCAGAGAGAAAAATAT 360
 QY 3559 taagttagcattctccctactcctaaagaataactgaatttatttgcctacctaagaat 3618
 Db 361 TAAGTAGTCATATTTCCCTACCTAAAGATGAATGATTTATTTGGTACCTAAGATG 420
 QY 3619 cagatatttagtttccattgcattgcatgctgtgtgtctatagacaatttcaatg 3678
 Db 421 CAGTATATTTAGTTTTCATTTGCAATGATGTGTGTGTATAGACAAATATTTAGATTG 480

RESULT 13
 AA990042/c 518 bp mRNA EST 02-JUN-1998
 LOCUS ua56f06.r1 Soares_thymus_2NDMT Mus musculus cDNA clone
 DEFINITION IMAGE:1361507 5' similar to WP.R11A8.4 CE06302 YEAST REGULATORY
 PROTEIN SIR2 LIKE ;, mRNA sequence.
 ACCESSION AA990042.1 GI:3175406
 VERSION EST.
 KEYWORDS house mouse.
 ORGANISM Mus musculus.
 SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 518)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Getzel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:699547
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28m13 rev2 ET from Amer sham
 High quality sequence stop: 219.
 Location/Qualifiers
 1..518
 location="Mus musculus"
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1361507"
 /clone_1lb="Soares_thymus_2NDMT"
 /sex="male"
 /tissue_type="thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGGTGGAGCGCGCGGTTTATTTTATTTTATTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento

BASE COUNT 144 a 100 c 98 g 174 t 2 others
 ORIGIN
 Query Match 12.4%; Score 478.4; DB 14; Length 518;
 Best Local Similarity 96.3%; Pred. No. 2.3e-95;
 Matches 499; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 1323 gacctccatgttatttgatctctctgaagatgagaccagtagactaattccaagt 1382
 Db 518 GACCTCCTCATTTGTANNATGATCTGCATGAAGTGAGCAGCAGTACACTAATTTCCAAGT 459
 QY 1383 tctatacccatgaagtgcct-caaatatataaataagggaacctttgcctactaca 1441
 Db 458 ACTATACCCCAATGAAGTGCCTACCAATATTTAAATAGGGAACCTGTGCATCATATACA 399
 QY 1442 ttctgattgtagagctcctcttgagagctgcgctgtctataatgaatgagttgtgcatagagct 1501
 Db 398 TTTTGATGTAGAGCTCTCTTGAGACTGCGATGTATTATTAATGACTTGTGTATAGCT 339
 QY 1502 aggttgatgaatgagcacaacttctgttaaccctgttaagagcttcagaaattactgaaa 1561
 Db 338 GGGTGGGATATGATGACCAACTTTGTGTATACCTGTAAAGCTTTGAGAAATTTACTGAAA 279
 QY 1562 acctcaagcccaaaaaagggaattgttcatctatacagagtgccacaacaccttca 1621
 Db 278 ACCCTCACGCCCAAAAGGAATTTGTTATATGAGATTTGCCAACACACCTCTTCA 219
 QY 1622 tattcggaagactcaagttcacctgaagaactgtgacacagagactctctgtgattgc 1681
 Db 218 TATTTGGGAAGACTCAAGTTTACCTGAAAGACTGTACCAAGACTCTTCTGTGATTGC 159
 QY 1682 tacactgttagaccaaagcaacaacaacaaatgttaatgattgaagatctgaaatcaag 1741
 Db 158 TACACTGTGTAGACCAAGCAACAACAACATGTTAATGATTGAGAGTATCTGATCAAG 99
 QY 1742 ttgtgtggaagaaaacacacagaagtagcagactagtaggaatgttgagaacattaat 1801
 Db 98 TTGTGTGGAGAAAGAAACCAACAGAGACTAGAGAAAGTGTGAGAAACATTAATGT 39
 QY 1802 ggaataaccagattttaagcgttgttgcactacg 1839
 Db 38 GGAATATCCAGATTTTAAGGCTGTGTGTTCCAGACTG 1

RESULT 14
 BE692325/c 552 bp mRNA EST 11-SEP-2000
 LOCUS uw10D12.x1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:3416255
 DEFINITION 3', mRNA sequence.
 ACCESSION BE692325
 VERSION BE692325.1 GI:10079949
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1092067
 High quality sequence stop: 461.
 Location/Qualifiers
 1..552
 location="Mus musculus"
 /organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3416255"
/clone_lib="Soares mouse 3nBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGAGGAGCGCGCCGCTGTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."

```

```

BASE COUNT      211 a      75 c      80 g      186 t
ORIGIN

```

```

Query Match      12.3% Score 476.8; DB 111; Length 552;
Best Local Similarity 99.2%; Pred. No. 5.3e-95;
Matches 490; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

OY 3343 ggaatcaatttcctcaactgcagaaactttagccttcaaacagcttcacccgcagaag 3402
|||||
DB 552 GGATCAATTTTCTCACTGAGAGAACTTTTACGCTTCAACAGTTCACACTCAGAAAG 493
|||||

OY 3403 tcagattatttcaagactcttggagacattggccccaatttaattcaatgtg 3462
|||||
DB 492 TCAGTATTATTATTAACACTCTTTGGACATTTGCCCAAAATTTAAATTCATGTGG 433
|||||

OY 3463 gtttagtatttatacaaaaaatgaattgaataatagctgttcttatagataaataac 3522
|||||
DB 432 GTTAGATTTATTATTAACAAAAATGATTTGAATATATAGCTTTCTTTATGCAATAAATAC 373
|||||

OY 3523 ccagttgagaccattactgcagagagagaagaattgaatagatcattccctactaa 3582
|||||
DB 372 CCAGTTAGACCACTTACTGCGAGAGAGAAAGTATTATGATGCTATTTCCCTACCT - 313
|||||

OY 3583 aaggttaactgaattattttagctacactaaagaatgcagatatttagtttccattgc 3642
|||||
DB 314 AAGTAACTGAAATTTATTTGGCTACACTAAAGAAATGCGATATTATTACTTTCCATTGCG 255
|||||

OY 3643 atgactgtctgtcgtatagacaatatatttaactgaaaaattgctttaattatttt 3702
|||||
DB 254 ATGATGTGTTTGTGCTATAGACAATATTTTAATTTGAAAAATTTGTTTAATTTATTTT 195
|||||

OY 3703 acagtgaaagactgttccagctctttatatagtacatagactttatgtaactggca 3762
|||||
DB 194 ACAATGAAGACTGTTTCAGCTCTTTTATATGTATGATATAAATCTTTATGTAATCTGCA 135
|||||

OY 3763 tatgttttgtagaccgtttaatagctgatatattctctccaacttttgaatacaaaa 3822
|||||
DB 134 TATGTTTGTAGACCGTTAATGACTGATATATCTTCCCAACTTTTGAATACAAAAA 75
|||||

OY 3823 cagcttttaact 3836
|||||
DB 74 CAGTCTTTTAAACT 61
|||||

```

```

RESULT 15
AA465858      479 bp      mRNA      EST      04-AUG-1997
LOCUS        ve88f01.g1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
DEFINITION   IMAGE:833305 5', mRNA sequence.
ACCESSION    AA465858
VERSION      AA465858.1 GI:2191998
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus

```

```

REFERENCE
AUTHORS
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 479)
Marras, M., Hallier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Teisling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
CONTACT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousestewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (InfoImage.lnl.gov) for further information.
MG:493521
Seq primer: -40ml3 fwd. ET from AmerSham
High quality sequence stop: 468.
FEATURES
Source
1..479
Location/Qualifiers
1..479
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:833305"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pBluescribe (modified);
Site 1: MluI; Site 2: SalI; cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
SalI(dT): 5'-CGGTCGACCGCGCAGCGTTTCTTTTCTTTT-3'. cDNAs
were cloned into the MluI/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."
BASE COUNT      143 a      74 c      76 g      186 t
ORIGIN

```

```

Query Match      12.1% Score 468; DB 7; Length 479;
Best Local Similarity 99.8%; Pred. No. 4.5e-93;
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 2288 aattgtccaccacattgggaactttagcatgtccaataaatagaatttacttgtaa 2347
|||||
DB 1 AATTGCTCCACACGACATTGGGAACTTTAGCATGTC-AAAAATGAAATGTTTACTTGTA 59
|||||

OY 2348 ctgacaagaagaatcgaagatgtattattatagaatggaataagatgtctctt 2407
|||||
DB 60 CTGACACAGGAATCTGAAGAATGATATATTATAGCTGGAATAATGATGTGCTTTT 119
|||||

OY 2408 ggaataattcctaagaatgccatcttctgttgactgtatcatcattcaacactgtgtg 2467
|||||
DB 120 GGATAATTTTAAAGTCCATCATCTTCTGTTGATCTGTACATTCACACACTGTGTTG 179
|||||

OY 2468 actcatcttctctcaagttcaatgtatgtatgatacatctgtatgataatatttgt 2527
|||||
DB 180 ACTTCATCTTCCTTCAAGTTCATTTGATGATATACATTCGTATGTATGTAATTTTGT 239
|||||

OY 2528 ttlttgctaagatttcaaccctttaaagtttcaaaaagcattggaatttaagt 2587
|||||
DB 240 TTTTTCCTATAGATTTCACACTTTTAAAGTTTCAAAAAGCCATTGGAATGTTAAATGA 299
|||||

OY 2588 aagggaacagcttactagacaaagaatggtatctcaacttttgttgttaacatgtg 2647
|||||
DB 300 AAGGAACAGCTATATAGACCAAGAATGATATTACACTTTTGTGTTGTAACATTCG 359
|||||

OY 2648 aatagtttaagccctcaattctgtctcgtgaactttatttggagacttaact 2707
|||||

```


Tue Feb 27 15:47:09 2001

Db 360 AATAGTTTAAGCCCTCAATTCTGTCTGTCGCTGAACCTTTATTTTATGAGACAGTTAACTT 419
Oy 2708 tttaaacacttggaattttccaaaacttftgagcttaactttttaaatcacagatgact 2767
Db 420 TTTAAACACTGGCATTTCCTCAAAACTTGTGGCAGCTAACTTTTAAATCAGATGACT 479

Search completed: February 27, 2001, 05:50:28
Job time: 42840 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:41:38 ; Search time 76.21 Seconds
(without alignments)
330.676 Million cell updates/sec

Title: US-09-461-580A-1

Perfect score: 3854
Sequence: 1 MADEVALALQAAGSPSAAAA.....NEAATRQELTDVNPSPDKS 737

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	479	12.4	257	21	Y73360
2	355.5	9.2	212	20	Y48540
3	138	3.6	205	19	W98718
4	135	3.5	1805	13	R27204
5	135	3.5	1805	13	R60126
6	132	3.4	1192	21	Y56967
7	122	3.2	738	19	W56163
8	121.5	3.2	1857	21	Y53970
9	121	3.1	531	20	Y40098
10	121	3.1	595	19	R14309
11	121	3.1	595	19	W53347
12	121	3.1	595	21	Y59071

13	120	3.1	873	19	W49031
14	119	3.1	180	13	R22392
15	117	3.0	465	19	W61003
16	115.5	3.0	597	20	Y07450
17	115.5	3.0	880	21	Y44638
18	114.5	3.0	1584	21	Y84430
19	114	3.0	1362	20	W99481
20	113.5	2.9	903	20	W99665
21	113.5	2.9	2559	20	Y41012
22	112	2.9	594	18	W15260
23	111.5	2.9	831	16	R80168
24	110.5	2.9	696	20	Y13359
25	110.5	2.9	696	21	Y70671
26	110.5	2.9	1786	13	R41043
27	110	2.9	2783	13	R23963
28	109.5	2.8	911	19	W74723
29	109	2.8	394	8	P70434
30	109	2.8	1166	20	Y08643
31	108.5	2.8	394	11	R04834
32	108	2.8	98	20	Y28479
33	108	2.8	98	20	Y25428
34	108	2.8	385	20	Y33672
35	107.5	2.8	527	20	Y29515
36	107	2.8	959	21	Y53051
37	107	2.8	1212	20	W88287
38	107	2.8	1212	20	W88288
39	107	2.8	1212	20	W88289
40	107	2.8	1212	20	W88290
41	107	2.8	1447	20	W81029
42	107	2.8	2441	19	W40058
43	106	2.8	330	11	R05528
44	106	2.8	1312	21	Y76884
45	106	2.8	2396	13	R29939

ALIGNMENTS

RESULT	ID	Description	Human Prt1-like su
1	Y73360	standard; Protein; 257 AA.	Antigen tc-7a. B1
XX	XX	XX	Streptococcus pneu
XX	XX	XX	Human TSI0q23.3 ge
XX	XX	XX	N. meningitidis T-
XX	XX	XX	Amino acid sequenc
XX	XX	XX	Murine p/cip prote
XX	XX	XX	Human secreted pro
XX	XX	XX	Amino acid sequenc
XX	XX	XX	Rat brain-specific
XX	XX	XX	PMISS1 Misp spider
XX	XX	XX	Amino acid sequenc
XX	XX	XX	Human PRO266 prote
XX	XX	XX	CD4-EBAl75 fusion
XX	XX	XX	AFP-1 (Ala 2460 Va
XX	XX	XX	Human secreted pro
XX	XX	XX	Amino acids 71-464
XX	XX	XX	S. aureus Sdrt pro
XX	XX	XX	N-myc protein. R
XX	XX	XX	Antifreeze peptide
XX	XX	XX	Synthetic antifree
XX	XX	XX	B. bassiana POPS O
XX	XX	XX	Human lung tumour
XX	XX	XX	Human secreted pro
XX	XX	XX	Rat (Na,K)-ATPase
XX	XX	XX	Rat (Na,K)-ATPase
XX	XX	XX	Rat (Na,K)-ATPase
XX	XX	XX	Murine PCIP protei
XX	XX	XX	Cellular transcrip
XX	XX	XX	High density lipop
XX	XX	XX	Retinoblastoma bln
XX	XX	XX	Deduced from lely

PT New peptides useful for diagnosis, prevention and treatment of cancer
 and immune disorders

PS Claim 1: Page 120-121; 193pp; English.

XX Y73325-Y73389 are human transcriptional regulator molecule (HTRM) protein
 CC sequences. The HTRM protein and nucleotide sequences are useful for
 CC preventing or treating disorders associated with decreased expression or
 CC activity of HTRM which include cell proliferative disorders such as
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and
 CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists
 CC of the HTRM polypeptides are useful for treating or preventing disorders
 CC associated with increased expression or activity of HTRMs. HTRM
 CC polypeptides, their immunogenic fragments or oligopeptides are useful for
 CC screening libraries of compounds in drug screening techniques.
 CC Polynucleotides encoding HTRM are useful for blocking the transcription
 CC of mRNA and regulating gene function by modulating the activity of HTRM.
 CC Vectors expressing HTRM or agonists can also be used to prevent or treat
 CC disorders associated with decreased HTRM expression. Antibodies which
 CC specifically bind HTRM and polynucleotides encoding HTRM are useful for
 CC diagnosis disorders associated with the expression of HTRM, particularly
 CC in assays that detect the expression of HTRM. Nucleotide sequences
 CC encoding HTRM may be useful to generate hybridization probes useful in
 CC mapping the naturally occurring genomic sequence and to detect
 CC differences in gene sequences among normal, carrier and affected
 CC individuals. Using diagnostic assays, cancer can be detected prior to the
 CC appearance of clinical symptoms and thereby progression of cancer can be
 CC prevented by aggressive treatment or preventive measures.

XX Sequence 257 AA;

Query Match 12.4%; Score 479; DB 21; Length 257;
 Best Local Similarity 41.9%; Pred. No. 1.8e-34;

Matches 103; Conservative 50; Mismatches 73; Indels 20; Gaps 7;

QY 251 LTVAGVSVSCGIPDERS-RDGIYARLAVDFPDPQAMFDIEYFRKDPPEFFAKETIY 309
 DB 1 mvgagstpsqjdftrpspsjysnl--qyglpypeafefpfhmpkpfllakely 58
 QY 310 PCQFQPSLCHKRTALSDEKGLRNTQNTIDLEOVAGI--ORLIQCHGSFATASCLICK 367
 DB 59 pgnypnvhyflrlhdkgllrltyqldglerysgipaaklveahgtfasatctvcg 118
 QY 368 YVDCNAVGDIFNCPVPCRCPADDEPLATMKPEIVFEGCMLPQGF--HRMKYDKDEV 425
 DB 119 rpfpedlradmadivpcpcvpc-----tgvvkdvlvtfgeplpqfllh---vvdfpma 170
 QY 426 DLLIVGSSLRVPVALIPSSIPHEVQQLINREPLPHLHF-----DVELLGCDDVITNE 480
 DB 171 dlllllgslvepfastleavrsvprlllnrdlvglawprdrdaqldvvhgves 230
 QY 481 LCHRG 486
 DB 231 lvelig 236

RESULT 2
 Y48540
 ID Y48540 standard; Protein: 212 AA.

XX Y48540;

XX 08-DEC-1999 (first entry)

XX Human breast tumour-associated protein 1.

XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
 KW treatment; tumour; cytostatic; medicament.

XX Homo sapiens.

XX DEL9813839-A1.

XX 23-SEP-1999.

XX 20-MAR-1998; 98DE-1013839.

XX 20-MAR-1998; 98DE-1013839.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-528981/45.

XX N-PSDB; 233613.

PT Human nucleic acid sequences and protein products from tumor breast
 PT tissue, useful for breast cancer therapy

PS Claim 25; 143; 188pp; German.

CC This invention describes novel human nucleic acid sequences from tumor
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer. Y48540-Y48617 represent protein fragments
 CC encoded by the expressed sequence tags described in the method of the
 CC invention.

XX Sequence 212 AA;

Query Match 9.2%; Score 355.5; DB 20; Length 212;
 Best Local Similarity 44.5%; Pred. No. 1.2e-23;

Matches 73; Conservative 35; Mismatches 49; Indels 7; Gaps 4;

QY 206 ELDDMTLMQIVINILSEPPKRRKKRDINTIEDAVKILQ--ECKKIIVLTGAGVSVSCGIP 263
 DB 42 eadmdflmlfsgtstlsqgkerllldeltlegvarymgsercrrylclvgagistcsaglp 101
 QY 264 DERS-RDGIYARLAVDFPDPQAMFDIEYFRKDPPEFFAKETIYPCQFQPSLCHKRTI 322
 DB 102 dfrstpsgtynhl--ekyllpypeafefsfkhppeffalakelypqgfkptichyfm 159
 QY 323 ALSDEKGLRNTQNTIDLEOVAGIQR--LIQCHGSFATASCL 364
 DB 160 rllkdkgllrlcytqldlerlaagledlveahgtfytshcv 203

RESULT 3
 W98718
 ID W98718 standard; Protein: 205 AA.

XX W98718;

XX 31-MAR-1999 (first entry)

XX H. pylori GHP0 808 protein.

XX GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.

XX Helicobacter pylori.

XX W09843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

```

DR      N-PSDB; Q283396.
XX
XX      Diagnosis of pre-disposition to brain tumours - using DNA
PT      encoding nestin which distinguishes neural multi-potential stem
PT      cells from neuronal, glial and muscle cells
XX
XX
PS      Claim 3; Fig 1; 63pp; English.
XX
CC      The amino acid sequence of the nestin gene which was deduced from
CC      the nucleotide sequence, suggests that nestin is a member of the
CC      intermediate filament protein family. The rat nestin amino acid
CC      sequence shows 75% similarity with the human nestin sequence. There
CC      is more than 60% identity between the two sequences. Antibodies to
CC      nestin protein can be used in in vivo diagnosis of brain tumours.
CC      See also Q28339 for the human nestin gene.
XX
XX      Sequence      1805 AA;
SQ
      Query Match      3.5%;      Score 135;      DB 13;      Length 1805;
      Best Local Similarity 19.7%;      Pred. No. 0.015;
      Matches 155;      Conservative 99;      Mismatches 262;      Indels 272;      Gaps 36;

QY      41 PGLGSPGEPSPAAYVAPAAAGCEAASAAAPALIMREAGAAASA---ERRAPA----- 89
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      392 p1isaacpcpnaevravagewp1sl1lqtgcawp1wk1katvysssa11pelaeppgkqghfp 451

```

Query Match	3.5%	Score 135	DB 13	Length 1805
Best Local Similarity	19.7%	Pred. No. 0.015		
Matches 155	Conservative 99	Mismatches 262	Indels 272	Gaps 36

QY	41	PELGSPGEPSAAVPAAPAGCEAASAAADALMREAAAGAASA	---EREAPA	-----	89
		: : :		:	
DB	392	ptiseacpcpnaevragewplslilqtgapeplwlkatypssalipelleepgkqgnfr			451
		: : :	: :		
QY	90	---TAVAGDGDNSGLRERPRADDPTFDGDGE	-----	-EEDEAAAAA	127
		: :	: :		
DB	452	daltslatnlmphptleakdgess	-----	earvsallfgedegqtwelveakadlevkve	506

```

OY 41 PELGASPGSPSAVAVPAAAGCEAAASAAALMREAAAGAAASA---ERFAPV----- 89
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 pRisaepcpPmaevraqevpRlIqltqapepIwIkatyPssalIpeIeapqkgqhtp 451
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 90 -----TAVAGDGDNGSGLRERPRADDPFDDGCE-----DEDEAAAA 127
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 ddltslatnIphhptleakdgess-----eervsalIgedeqIwelveakdeIvke 506
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 128 AAAAIGYRDLNLTGGLLTNGFHCESDDDDRTSHASSSDMPPRIRIGYTVQOHLM-- 185
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 nsaadktetgsjdlte-----etqdsq-----gp---lqketIka 537
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 185 TATDDEETIIVKLTIDMPTDDEDTDDMTYVATGATITVYKTLICENQVKK-- 235
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

```

Db      507  nsaaqktqdsqjldte-----etqdsq-----gp---lqkeltla 537
      508  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      166  iSTDPTILIKDLPEETIPPELDMMTLMOIVINIISEPKR-----KRDI---NTI 235
      509  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      538  lgeepImslkIqnyetbaqkencnsat--eqhgtIegpekeqjDlksIeenvesekfl 595
      509  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      236  EDAAVLLOECCRKIIVLTGAGVSVCSGIPFRSRDGIYALADPDLPPQAMFIEYPR 295
      510  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      596  enqyprylsI-----lqkredtrcd-----qelmspyktl----- 625
      510  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      296  KDPRPEFKAEIYQGQFQPSLCHKFIALSDKEGKILLRNYTONIDTLEOVAGIQRILOCH 355
      511  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      538  lgeepImSkIqnyebaqkencnsat--eghIgtIegpekexqJlksIeenvesekfL 595

Oy      236  EDaVnLlOeCCKKIIVLTGAGvSvSGiPFRSRDGiYARLADPDLpPQAMFdiEYPR 295
      596  enyvpyIsel-----Iqkedrtred-----qelmspxgtl----- 625

Oy      296  KDRPFPEFAKEIYvGQvQPSlCHKFIAlSDKEGKLLRNyTONIDTLEovAGiQRILOCh 355
      626  ---kTfssIyge-----sqevyrpskegnIeswtafkeesqnpIlg----- 662

Oy      356  GSEfATASclICkYKvDCeAvRGdIRINovvPRDPRADpELIMKPEIvFFeENLPEQFH 415

```

```
Db      596 enyvnylsel-----lqkredtrede-----qelmssygtl-----        625
QY      296 KDRPPEFKAKEIYPGQEPQLSCHKMFIALSDKEGKLLRNYNTONIDTLEOVAGIQRILOCH    355
           |         |         |         |||         |         |         |
Db      626 ---kfisslygk-----sqevnrypskegn--lewtfkessqprlg-----        662
QY      356 GSEFATASCLCKKKVYDCAEVRGDITRNQVVPRRCRPADRLPLIMKPFIYFFENLPEOPH    415
           |         |         |         |         |         |         |
Db      663 --fpgaaedmlerlveked-----qsfrtspreeedgaacrp-----lqkenq     702
QY      416 RAMKYDKDDE---VDLLIYGSSLKRVPAVLIPSSIRPIHEVPOL--INREPLPHLHFDEL     470
           |         |         |         |         |         |         |
```

```

Db      626  --krtsslyge-----sgewvrpskegn--leawtafkreesqnp1g----- 662
QY      356  GSEFATASCLICKKYKDCAEVRGDIINQVVRPRRCRADPRLAIMPELVFFENLEPEQN 415
Db      663  --fpaedqmler1evked-----qsfrrispeedgeacpr-----1qkenq 702
QY      416  RAMKVDKDE--VDLLIVGSSLKRPVALIPSSIPHEVPQTL--INREPLRLHFDVLL 470
Db      703  epllyeeseqqler1lekesqsrls---peedgeqsrlqkenqnp1lyeae--- 755
QY      471  LGDCVYIINELCHRIIGGEYAKLICCPVKISSETTEKPRP--QKELYHLSLPRP----- 522

```

```

Db      663  ---lpgaequmlerlveked-----qslrpsrpeeqgaacrp-----lqkeng 702
Qy      416  RAMKDKDE---VDLLIVGSSLKRPVALIISSTPHEVPQL--INREPLHLHFVDL 470
Db      703  eplgyeeaeqqlerllekesqslrs---peedeqeagrlqkengpplgyeeae--- 755
Qy      471  LGDCVYIINELCHRIIGGEYAKLCCLCNPKVKSITETKPRP-QKELYHLSLPT----- 522
Db      756  ----qqlerllekesqslk---speeqrlgkrlatengqsllyleengetfypleær 808
Qy      523  -----PLH-----ISDSSSPERVPQDSSVIATLVQDATNN 554
Db      809  nrrrlrlveeeearrrvrllekrcdsrlaeeenwvrlyleeaddcrlnkrllektrk 868

```

[illegible]

```

0Y 523 -----PLH-----ISESSSPEPTVPODSSVIALTVDOATNN 554
Db 809 ngrplrslsveeeegqrvkvpklyeqsgdsygsaaenvaplrlyleeddcinksllektlk 868
0Y 555 NVNDLE-----VSESSCSEERKPOEVOTSRNENINVENPDKFAVGSSTADKNER 603
Db 869 slyslsiedrngdsllipgeetqvalrpeedeqqrlvnhleksegef-----sr 916
0Y 604 TSVAEIVRKCMWNLAKESIKR-LEGNGLYLVPVNNRYFHCFAEYVSQSEDDPVLSSSCG 662
Db 917 ssee-----eqymerslegen-----hslslysvkedqmvveqlek 953

```

QY 663 S N S D S G T C Q S P L E E P L E D E S E I E E F I N G L E D B T E R P E C A G G S G F G A D G S G D E V N E A I A 722
: ||| : |||| : | : | : |||

```

Db      954 esqdgsg-----ksledes--gctfgleknae-----slrslagdqdeqklege .997
Oy      723 TRQELTDV 730
        1:1111
Db      998 lqqlrrav 1005

RESULT      5
R60126
ID      R60126 standard; Protein; 1805 AA.
AC
XX      R60126;
XX
DT      21-MAR-1995 (first entry)
XX
DE      Rat nestin protein is useful to identify brain tumours.
XX
KW      nestin gene; brain tumour; neoplastic cells; glial; neuronal;
KW      muscle; neural multipotential stem cell; mammalian brain; detection;
KW      diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.
XX
OS      Rattus rattus.
XX
XX      Key Location/Qualifiers
FH      Misc-difference 167..168
FT      /note="encoded by CACGC"
FT      Misc-difference 172..173
FT      /note="encoded by CCG G GCC"
XX
XX      US5338839-A.
XX
XX      16-AUG-1994.
XX
XX      12-APR-1988; 88US-0180548.
XX
XX      12-APR-1988; 88US-0180548.
XX
XX      02-JUN-1988; 88US-0201762.
XX      02-JUN-1988; 88US-0201762.
XX      25-OCT-1990; 90US-0603803.
XX      25-OCT-1990; 90US-0603803.
XX      22-FEB-1991; 91US-0660412.
XX      22-FEB-1991; 91US-0660412.
XX      19-MAR-1992; 92US-0853913.
XX
XX      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX      Lendahl U, McKay RDG;
XX
XX      WPI: 1994-263332/32.
XX      N-PSDB: Q70447.
XX
XX      Nucleotide and protein sequences for human and rat nestin -
XX      distinguishes neural multipotential stem cells and brain tumour
XX      cells from more differentiated cell types; for use in the
XX      diagnosis of brain tumours
XX
XX      Claim 4; Column 35-44; 45pp; English.
XX
XX      This sequence is the rat nestin protein encoded by Q70447. Its
XX      degree of sequence homology to intermediate filaments in the core
XX      domain, its dissimilarity in the head and tail domains, and its
XX      different splicing pattern suggest that it defines a new class of
XX      intermediate filament protein. Nestin protein expression distinguishes
XX      neural multipotential stem cells and brain tumour cells from the more
XX      differentiated neural cell types (eg., neuronal, glial and muscle
XX      cells of the adult brain). The nestin protein can be used in diagnosing
XX      tumours of the brain, such as medulloblastomas, glioblastomas and
XX      oligodendroglioma. (See also R60127).
XX
XX      Sequence 1805 AA;
XX

```

Query Match	3.5%	Score 135;	DB 15;	Length 1805;
Best Local Similarity	19.7%	Pred. NO. 0.015;		
Matches 155;	Conservative 99;	Mismatches 262;	Indels 272;	Gaps 36;

OY	41	PGGSGSEEPSEAAVAPAAAGCEAASAAAPALMEAGAASA	---CREAPA-----	89	
Db	332	PLPSeapcpnpnaevragervp	sl qtqaperpwlwtkatvpsstallpelcepgkqgghfr	451	
OY	90	-----TAVAGDGDNGSGLRARRPRAADFDDECE	-----EEDEAANA	127	
Db	452	ddltslatnlnpnhpntlekkages	-----esrsvsllfgedegglwelvekeadievke	506	
OY	128	AAAIAGYDNLITLDGLITNGFHSCESDDDRTSHASSDWT	PPRIGPYTEVOQHLM--	185	
Db	507	nssaqktqesglde	-----etqdsq-----gp--lqketlka	537	
OY	186	IGTDRITLLKDLLEETIRPPRLDMLQIYINLSEPPRKK	-----KRKDI---NTI	235	
Db	538	lgeelmisklqnyetagekencnst	eghllgtlegpekqipklsleeknvesektl	595	
OY	236	EDAVALLQECKRIIVLTGAGSVSGIIPDFRSRPGIYARLAVDP	PPDIPDPOAMPDIEYFR	295	
Db	596	engvpviseel	-----lqkedtted-----qelmspkgtl	625	
OY	296	KDPRFFKFAKEIYPGQFQPSLCHKFIALSDKEGKLLRNYQ	NTIDTLEOVAIGIORILOCH	355	
Db	626	---krfsalgr	-----sqevrnpkqegn--leswtafkreesqhrpg	662	
OY	356	GSFATASCLIKKIVKDCAVAVGDIIFNQVVPKCPKPAD	EPALMKPELIVFEGGELNLPQFH	415	
Db	663	---fpnaedqmlerlvekcd	-----qsftrpspeedqeaocrp-----lqkeng	702	
OY	416	RAMKXDKKE	---VDLLIYIGSSLKVRPVALIPSSILPHEVPIIL--	INREPLPHLFPVEL	470
Db	703	eplygeeeqglterlllekesqeslrs	-----peeedeagrsllkengdeplygeee	755	
OY	471	LGDDVITINELCHRLGGEYARLCCNPYKLSITEKPRPR	-OKELVHLSELPT-----	532	
Db	756	---dqmterlllekesqeslk	---speengrlygkprlcrendqkslryleengqetfvplesr	808	
OY	523	-----PLH-----	ISDSSSPERTVYQDSSVATILVDQATNN	554	
Db	809	nqrplrsleveeeeqrlvplekysqdsllgslaeenqplrlyleedc	llksllsedcthk	868	
OY	555	NWNIDE	-----VSESSGVEEKPOEYQOTSRNENINVENPDRKAVGSS	STADKNER	603
Db	869	slgsledtngdslllpgeesecq	vsllrppceedgrlyvnhlekesqel	-----str	916
OY	604	TSVAETVRKCPMNLRAEQISKR	-LEGNOYLEVPBNRYTFHGAEVYSDESDVYLSSSSCG	662	
Db	917	ssee	-----eqvmerslegen-----heslsvvekedqmwesglek	953	
OY	663	SNSDSGTSQSSLEPPLDESEIEEFYNGLEDOTERBPCAGSG	SFGADGQDEVYNAIA	722	
Db	954	esqdsq	-----kslides-qetifplekenae-----slrsiaqdgdeeqklege	997	
OY	723	TROELTDV	730		
Db	998	tqqltrav	1005		
RESULT	6				
ID	Y56967				
AC	Y56967	standard; protein; 1192 AA.			
XX	25-APR-2000	(first entry)			
DE		Human MAGI polypeptide.			
KX		MAGI protein; neuroendocrine-specific protein; neuropathy; human;			
KW		spinal injury; neuronal degeneration; neuromuscular disorder; cancer;			
KM		psychiatric disorder; developmental disorder; inflammatory disorder;			
KW		stroke; cytostatic; cerebroprotective; neuroprotective.			
OS		Homo sapiens.			

XX PN WO200005364-A1.
 XX PD 03-FEB-2000.
 XX PF 21-JUL-1999; 99WO-GB02360.
 XX PR 22-JUL-1998; 98GB-0016024.
 XX PR 19-JUL-1999; 99GB-0016898.
 XX PA (SMK) SMITHKLINE BEECHAM PLC.
 XX PI Michalovich D, Prinjha RK;
 XX DR WPI: 2000-182693/16.
 XX DR N-PSDB: 256886.
 XX PT Novel polypeptides related to neuroendocrine-specific proteins and
 PT polynucleotides useful for diagnosis of various diseases and for
 PT treatment of cancer and neurological disorders -
 PS Claim 2: Page 20-21; 35pp; English.
 CC The invention relates to human MAGI protein, which is similar to
 CC neuroendocrine-specific protein. The MAGI protein can be expressed by
 CC standard recombinant methodology. The MAGI polypeptides, polynucleotides
 CC and antibodies are useful for treating diseases, including neuropathies,
 CC spinal injury, neuronal degeneration, neuromuscular disorders,
 CC psychiatric disorders and developmental disorders, cancer, stroke and
 CC inflammatory disorders. The polynucleotide is also useful for chromosome
 CC localization and for tissue expression studies. The present sequence
 CC represents the human MAGI protein.
 XX
 SQ Sequence 1192 AA;

Query Match 3.4%; Score 132; DB 21; Length 1192;
 Best Local Similarity 18.9%; Pred. No. 0.015;
 Matches 176; Conservative 110; Mismatches 356; Indels 290; Gaps 38;

QY 12 AGSPSAAAMEASQPADEPLRKRRRDPOLGRSPGEPSAANVAPAA-----AGCEAA 64
 DB 119 aaplsaaavspklpdeddeparp-----pppppsavspqgepwtpapapaa 168
 QY 65 SAAPALALMREAGAA-----ASAE---REAPATAV-AGCDNNGSGL 102
 DB 169 ppscpaapkrsgsvdelalfalpaasepyisaaemdlkeqgnltisagqdfpsvl 228
 QY 103 RR-----EPRADDFDDEGEDEEAAAAAAGYNDNLTLDTGLTNGFHSCE 154
 DB 229 letaasipsipisaasfkehe-----yignlstvlpcegligenvseask 274
 QY 155 D-----DDRTSHAS-----SSDWTPRPRIQPYTFVQOHLMTGTPR--TTLKD 196
 DB 275 evsekakllldldlftseleysemgsfsfspk-----aasavlyanpreiellvkn 327
 QY 197 -----LLPETI-----PPPELDMTLQIVIN-----ILSEPPAKKKKKDI 232
 DB 328 kdeeekivsnmllhngqelptalklkivkedevasekakdsfnekryaveagmreeyadf 387
 QY 233 NTIEDAVKLLQOECKIIVLNGAGSVSGGIPD-----FRSRDGIARLAVID 278
 DB 388 kpte-rwewkdskedmdlaagklesnleskvdckcfadslqcthekssnddts 446
 QY 279 FPDLPD-----PQAMPDIEFRKDPPEFFKAKEIYPGQFQPSLCHK----- 320
 DB 447 fspesgikdrpsayitcapf--npatesiatniflpgtbsentkdekkieekaqi 504
 QY 321 -----FILSDKEGKLRLNYTONID--TLEOVAG-----IQRIILQCH 355
 DB 505 vteknstksnplvaagqsetdyvt--tdnlkvtleevevanmpgltpdlvgacese 562
 QY 356 GSFAATSLCKIKYKVDCEAVAGDIFNOVVPKRCPCPADE-----PLAIMKREIYF---FG 407

DB 563 lnevtgltkiayetkmdlvqisevmqgeslypaaglcpsfseaseapspvyl-pdlvmeapln 621
 QY 408 ENLPEQFHRAMKDYKDEVDLLIVIGSSILKRVAPVLLP-----STIPHEVQIL 455
 DB 622 savpsagasyvlpsspslpaassvnyesikhepenppyeaamsvlykkvsglikeel---- 677
 QY 456 INREPLHLHFDVLLDDCDVITINELCHRLGGEYAKLCQCNVKSISEIT----- 503
 DB 678 --kep-----enlnaalgeteapylisacdliketklsaeapadfsdys 720
 QY 504 -----EKPPRPQKELVHSELPTPLHISEDSSPERYPODSVIATLVQDATNNVNDL 559
 DB 721 makveqpydhiselvedsspsdsepvdlfsdssldp--vpqgqdetmvlvkeslc----- 772
 QY 560 EVSESCVEEKPKQEVQTSRVNENINVENPDKAVAGSSTADKNERTSVAETVRKQWPNRLA 619
 DB 773 etsfsmleyenkelsalpppegkpylesfklsldntkdlldpdevs-tlskkekiplq 831
 QY 620 KEQISKRLGEGQVLFVPPNRYIFHGAEYSDS-----EDDVLSSSSCGS----- 663
 DB 832 meelstavsnldlfliskeagi-retelfdsdpelldetpctllsktdsfklareyt 890
 QY 664 -----NSDSGTQCSPSLEPLE-----DESEIEEYNGLEDTERPECAGSGSF 707
 DB 891 dlevshkselanapdgagslpctelpndlsiknlpkveekistsdd-----f 938
 QY 708 GADGQDEVV-----NEAINTQGLTDVNP 733
 DB 939 skngsatskvllppdvsalatgaefiesivkp 970

RESULT 7
 ID W56163 standard; Protein; 738 AA.
 XX
 AC W56163;
 XX
 DT 28-JUL-1998 (first entry)
 XX
 DE New DNA sequence isolated from Pinctada fucata.
 XX
 KW Pinctada fucata; protein; cosmetic.
 XX
 OS Pinctada fucata.
 XX
 PN JP10080285-A.
 XX
 PD 31-MAR-1998.
 XX
 PF 28-MAY-1997; 97JP-0138461.
 XX
 PR 15-JUL-1996; 96JP-0184459.
 XX
 PA (MIKI-) MIKIMOTO SEIYAKU KK.
 XX
 DR WPI: 1998-254410/23.
 DR N-PSDB: V22683.
 XX
 PT New cDNA and e.g. vector, host cell and polypeptide - used to
 PT produce polypeptide in high yields, which is used in cosmetics
 XX
 PS Claim 9; Pages 9-11; 15pp; Japanese.
 XX
 CC The present sequence represents protein encoded by a new DNA sequence
 CC isolated from Pinctada fucata. The protein be used as an ingredient
 CC in cosmetics.
 XX
 SQ Sequence 738 AA;

Query Match 3.2%; Score 122; DB 19; Length 738;
 Best Local Similarity 32.3%; Pred. No. 0.053;

Matches 61; Conservative 9; Mismatches 65; Indels 54; Gaps 8;

OY 5 VALALQAG-----SPSAAMEASQAPDEPLKRRPRDGPGLGR 45
Db 461 laaalaagaagggfgyglgglggsaaaaaaagggatrralrrlrrmr99gs 540
OY 46 SPGEPSAAVAPA-----AAGCEAASAAAPALMREAGAAA 81
Db 541 aaaaaaaagggwggmgggfgyglg9gf9ggf9gssaaaaaa-----aagf9g 596
OY 82 SAERAPATAVAGDDNGSLRRERPADDFDDDEGEEDFAAAAAAATGTRNLILT 141
Db 597 gg-rrgrgrgrg9gdg-gngasavaaaaaaaggsaadvaaaaaaam-ygdg---a 650
OY 142 DGL-LTNGF 149
Db 651 dgpdtfngf 659

RESULT 8
ID Y53970 standard; Protein; 1857 AA.
AC Y53970;
DT 13-MAR-2000 (first entry)
DE Human peripheral benzodiazepine receptor associated protein-1.
XX KW Human; peripheral benzodiazepine receptor associated protein-1;
KW PRX-1; peripheral benzodiazepine receptor; chromosome 17;
KW central nervous system; immune system; gene therapy;
KW PRX-1 deficiency condition; endocrine system.
XX OS Homo sapiens.
XX PN .W09960117-A2.
XX PD 25-NOV-1999.
XX PF 06-MAY-1999; 99WO-FR01070.
XX PR 15-MAY-1998; 98FR-0006190.
XX PA (SNFI) SANOFI-SYNTHELABO.
XX PI Casellas P, Gallegue S, Jbilo O, Le Fur G;
XX DR WPI; 2000-062455/05.
XX DR N-PSDB; 236989.
XX PT New PRX-1 polypeptide that interact with peripheral benzodiazepine
PT receptor, used to treat e.g. Immune, central nervous or endocrine
PT disorders
PS Claim 1: Page 39-44; 44pp; French.
XX CC The present sequence represents a peripheral benzodiazepine receptor
CC associated protein-1, designated PRX-1. PRX-1 interacts specifically
CC with the peripheral benzodiazepine receptor. The PRX-1 gene is
CC localised on chromosome 17 in the q22-q23 region. The gene is
CC associated with markers of pathologies of the central nervous system
CC or immune system. The PRX-1 nucleic acid is useful in gene therapy (of
CC PRX-1 deficiency conditions, e.g. disorders of the central nervous,
CC immune or endocrine systems; as a source of diagnostic primers and
CC probes (see 236990-237023) and of antisense therapeutics; for
CC recombinant production of the PRX-1 protein; and for detecting allelic
CC variants, mutations, deletions, insertions, loss of heterozygosity and
CC gene rearrangements in the PRX-1 gene. The PRX-1 protein is used to
CC raise specific antibodies and to screen for specific modulators
CC (potential therapeutic agents). The antibodies are used as immunoassay
CC reagents, e.g. for diagnosis of abnormal expression or accumulation of
CC PRX-1.

XX SQ Sequence 1857 AA;
Query Match 3.2%; Score 121.5; DB 21; Length 1857;
Best Local Similarity 19.3%; Pred. No. 0.23;
Matches 171; Conservative 82; Mismatches 233; Indels 381; Gaps 46;

OY 2 ADEVALALQAGSPSAAMEASQAPDEPLKRRP---RRDGP----- 42
Db 1067 adsiapitpalapslparyscpsphspearaplaaspgygdpspplqhpelqtce 1126
OY 43 -LGRSGEPSAAVA-----PAAAGCEAASA-----AADAA 71
Db 1127 ppgappaspremakysnedppapcsgeaagavlytseartastllygkdp9paapsl 1186
OY 72 LMREA---AGAAASAE-----REAPATAVAGGDNGSLRRERPA----- 108
Db 1187 akgeaewtlaageacpassstgaraqapntemc9g9dps9glr--praekedtaelgyhl 1244
OY 109 -----ADDFDDEGEEDFAAAAAAATGTRNLILTDLTNGFHS----- 151
Db 1245 vnslydnginsldsdigeeeeeeeeeejgstrcsfkgq-vagnslrengakspqpf 1303
OY 152 CESDDDDRT-----SHASSD-WTPRR 173
Db 1304 cetdsdeellqgllelpdqfcskklfsipeeeeeeekekegagcsrdpp9pepa 1363
OY 174 IGPYTFVOQHLMTGD---PRITLKDLPETIP-PPE-----LDDMTLMQIVNLT-- 220
Db 1364 I-----lglgcds9qpr-----pgqclspessr9agdcledmp-----gl9vg 1402
OY 221 -----SEPPKRRKRDINTIEDAVKLLQEFCKITIVLGAVSVC9GIPDRSDG 270
Db 1403 ssrrtggsgspekprsttrppdr-----ehsrll--lsmngqasgrl9pre9g 1451
OY 271 IYARLAVDFPDLDPQAMFIE-YFRKDPREFFAKEIYPOGFQSLCKHFIALS---D 326
Db 1452 I-----pvlegprtlgaesgrl9psrrcsgrale9p-lascslpklelsie9d 1502
OY 327 KEGKLLRNTQNTIDTLEQVAGIQRLQCHGFATASCLICKKVVCEA-----VAGD 378
Db 1503 sed-----eqeag-----sg9lsitssc-----ypgdrea9glatv9r9p 1539
OY 379 IFNQVPR-CPRCPADEPLAIKPELVEFGENLPPOFHR-AKKYKDEVDLITVGS-- 434
Db 1540 pkansgpkpyrtilpawek-----ge--petgrsatgrakepistratetge9g 1586
OY 435 -----LKVRYVALIPSSIPHEVPOILINREPLPHLHFDVL---LGDGD 475
Db 1587 qdgsgrt9pqrk9rvrvl9pstaetl9vars9et-----layqhl9rvl9fal9dyd 1637
OY 476 VII-----NELCHRL-----G9TYALCCNPVKLSITEKP 506
Db 1638 pvsmspn9dageeelp9f9egllkvfgdkdad9fy9ge9g9rtyl9pcmv--a9av9ds 1695
OY 507 PRPOKEIV-----HLESLPPTPLHISE-PSSSPERTVPQDS 541
Db 1696 p9grt9ql9r9yl9pdll9l9egs9ng9r9f9y9st9ht9tp9p9k9rr9sk9ae9g9p9cp9pp 1755
OY 542 SVIATLVDOATNNVNDLEVSSES-SCVEEKPOEVQTSRNV----- 581
Db 1756 k9vp9a-----dlk9p9sm9wa9ald9np9q9ss9pmd9vea9elp9f9ag9dl9v9fg9ld 1805
OY 582 -----NIN-----VENPDFKAVGSSTADKNERT9SAEIVR 611
Db 1806 dd9fy9gelng9r9l9vp9n9f9eg9p9e9g---ld9re9pt9p9ae9qr 1849

RESULT 9
ID Y40098
XX Y40098 standard; protein; 531 AA.

XX

PI Reid RH, Zarfos PN;
vxy

DR WPI: 1998-286586/25.
DR N-PSDB: V37391.

PT New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT e.g. for identifying anti-bacterial(s) for treatment and prevention
PT of meningitis

XX
PS Claim 11; Page 98; 130pp; English.

XX
CC The sequence is that of the polypeptide encoded by a region isolated
CC from S. pneumoniae. The protein, or agonists of it,
CC may be useful as an antibacterial for treatment or
CC prevention of infection, specifically caused by S. pneumoniae
CC (particularly meningitis) but possibly also Helicobacter
CC pylori (ulcers and gastric cancer). It may be of particular
CC use before insertion of an in-dwelling device or any other
CC invasive procedure. The protein, or nucleic acid encoding
CC it, can also be used in vaccines to induce a cellular
CC and/or humoral immune response, or to screen for other
CC antibacterials. The DNA may also contain flanking sequences
CC that are potential sources of control elements for bacterial
CC gene expression. Detecting a sequence encoding the protein
CC can be used diagnostically, e.g. to detect a mutation for
CC serotyping or classifying infectious agents.

XX
SQ Sequence 465 AA;

Query Match 3.0%; Score 117; DB 19; Length 465;

Best Local Similarity 21.3%; Pred. No. 0.071;
Matches 69; Conservative 44; Mismatches 113; Indels 98; Gaps 14;

QY 381 NQV-----PRCRCPADPELAIMPEIVFEGENTPEQFRAMKYDKEDVLLIV 430
DB 59 ngveteleapkeapkteespksevkpt---dtlpkv--eegkedsaepsvee 112
QY 431 IGSSLKVR--VALIPSSIPHEVPQILINREPLPHLHFDVLLGDCDVIINELCHRLGG 487
DB 113 vgeveskpeekvavkpesqdkpaeskev----- 145
QY 488 EYAKLCNPVKLSITEKPPRQKELVHSELPTPLHISEDSSPERTVQDSSVIATL 547
DB 146 -----pveqakvpqbpvqptq-----aegpslp---kessqgen--pkedrgaet 186
QY 548 VDOATNNNNVNDLEVESSESCVEEKPOEVQTSRVENINVENPDKAVGSSTAD-KNERTSV 606
DB 187 pkgedegpaeeqelkveepveesktv--npveqpkvetpavkqtepteekvevtsi 244
QY 607 AETVKKCPNRIAKEO-----ISKRLGNGQYLFPPNRYIFHGAEVYSDSEDDVLSSSSC 661
DB 245 pqttr--yeedltkehgtrrevkngsrvtlp---ylina----- 282
QY 662 GSNSDGTGQSPSLPEPLEDESEI 685
DB 283 ---tgdttegtsttdeaeakev 303

Search completed: February 16, 2001, 10:43:12
Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:41:39 ; Search time 80.58 Seconds
(without alignments)
621.032 Million cell updates/sec

Title: US-09-461-580A-1
Perfect score: 3854
Sequence: 1 MADEVALALQAAGSPSAAA.....NEAATROELTDVNPSPDKS 737

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR66:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	923.5	24.0	607	T24172	hypothetical prote
2	621	16.1	670	S36616	regulatory protein
3	605.5	15.7	471	T39571	probable regulator
4	592	15.4	503	S59698	HST1 protein - yea
5	568	14.7	562	1 RGBYS2	regulatory protein
6	538	14.0	331	T18247	transcription regu
7	513.5	13.3	332	T40929	transcription regu
8	484	12.6	381	UC4639	silent information
9	439	11.4	357	S59678	HST2 protein - yea
10	349	9.1	247	H69827	conserved hypochet
11	346	9.0	246	A72370	regulatory protein
12	341.5	8.9	447	S54631	HST3 protein - yea
13	330.5	8.6	253	H69263	transcription regu
14	295	7.7	245	D29659	transcription regu
15	285.5	7.4	247	D72562	hypothetical prote
16	282.5	7.3	249	F71085	hypothetical prote
17	267.5	6.9	250	C75101	transcription regu
18	267	6.9	250	A83506	probable cobalamin
19	256	6.6	415	T50106	hscap [imported] -
20	245	6.4	239	C70486	conserved hypochet
21	243.5	6.3	370	S52699	hypothetical prote
22	220	5.7	287	T22324	hypothetical prote
23	211	5.5	237	H70554	hypothetical prote
24	210	5.4	259	E82191	probable nicotinat
25	210	5.4	279	E64656	hypothetical prote
26	200.5	5.2	236	A83495	hypothetical prote
27	197.5	5.1	234	A71838	hypothetical prote
28	190	4.9	287	T22325	hypothetical prote
29	187.5	4.9	233	G81307	probable transfera

30	186	4.8	3534	2	T42567	legument protein 2
31	185.5	4.8	246	2	G75570	conserved hypochet
32	179	4.6	1095	2	T25520	hypothetical prote
33	158.5	4.1	190	2	T35951	probable STR2 fam1
34	150.5	3.9	3421	1	WZBBB6	367K legument prot
35	147.5	3.8	460	2	T33110	hypothetical prote
36	146	3.8	118	2	T46348	hypothetical prote
37	146	3.8	2484	2	T26216	hypothetical prote
38	146	3.8	2607	2	T26215	hypothetical prote
39	143.5	3.7	2639	2	T31328	hypothetical prote
40	142	3.7	1137	2	G70868	fibroin - Chinese
41	138	3.6	205	2	C64677	probable regulator
42	135	3.5	1805	2	A34736	conserved hypochet
43	134.5	3.5	587	2	B70884	nestin - rat
44	134	3.5	1857	2	T50513	hypothetical prote
45	133	3.5	1906	1	S68235	hypothetical prote
						myosin-II light-chain

ALIGNMENTS

RESULT	1
T24172	hypothetical protein R11A8.4 - Caenorhabditis elegans
C/Species:	Caenorhabditis elegans
C/Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession:	T24172
R/Cummings, P.	submitted to the EMBL Data Library, March 1996
A:Reference number:	Z19849
A:Accession:	T24172
A:Status:	preliminary; translated from GB/EMBL/DBD
A:Molecule type:	DNA
A:Residues:	1-607 <MIL>
A:Cross-references:	EMBL:Z70310; PIDN:CAA94364.1; GSPDB:GN00022; CESP:R11A8.4
C:Experimental source:	clone R11A8
C:Genetics:	
A:Gene:	CESP:R11A8.4
A:Map position:	4
A:Introns:	18/3; 68/3; 111/3; 206/3; 500/3; 565/3
Query Match	24.0%; Score 923.5; DB 2; Length 607;
Best Local Similarity	36.5%; Pred. No. 1.5e-50;
Matches 217; Conservative 106; Mismatches 186; Indels 85; Gaps 15;	
Oy	158 DRTSHASSDWTTPRR-IGPYTFOQHLMIGTDPRTIKDLPE--TTPPELDMTLMQ 214
Db	55 ESTTSVSESQNNDEMSNRRQRLLDGCATPLQIIQIFPFMNSRIATNSEMHFA 114
Oy	215 IVINIILSEPPRRKKRKDINTEDAVKLLQECKTIIVLTGAGVSVCGLPDRSRDGIYAR 274
Db	115 ILSLLETRAPVRKLTFTVNSLADAVELFKTKHILVLTGAGVSVCGLPDRSRDGIYAR 174
Oy	275 LAVPEPLPDQAMEDIYEPKDRPEPKAKETIYPOFQPSLCHKFTALSDKRGKILRN 334
Db	175 LRSEFPDLPDPTAFMDYRERENPAPYNEARELFPQFVPSVSHRFTKELETSGRLRN 234
Oy	335 YTGQIDTLEQVAGIQRILQCHGFATASCLCKKVKVCEAVRGDIFNQVPRCPRADE 394
Db	235 YTGQIDTLEHQTGKRYVECHGSFKCTCGKGYGNGENREVLAMVAHCKRCE--- 291
Oy	395 PLAIKPEIYFEGNLPDQFHRAMKYKDEVLLIVIGSSLKVRVALIPSSIPHEVPOI 454
Db	292 --GVIKPNIVFFGDLGREFHQHTEDKKHVDLIVIGSSLKVRVALIPCHVCKKNPOI 349
Oy	455 LINEPPLHLFDVLLGDDVITINELCHRGEXATLCCPVLTSE----ILEKPRP 509
Db	350 LINESLPHYVADTELLGNCDDIIRDCISLGGSFTELITSYDSIMQCGTKSKQKPSQN 409
Oy	510 QKELVHSELPEPTPLHT-----SEDSSEPRTVQDSSVITATLVDAQTNNVNDLEVES 564
Db	410 KROLISQEDF---LNTCKMEKRRNDSSDEPTLAKPRMSVA---DDSMDSKKNP----- 457

Oy 565 SCVEEKPOEVOTSRNVENINVENPDKFAGVSSSTADKNERTSVATEVRKCKWPNRLAKEOIS 624
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 458 -----QEIQKHKSEDDDDDTFNSD-----DLIKKIKHRL--ST 490
 Oy 625 KLEGNQYLFVPPNRYIFGAE-----VYSDSEDDVSSSCGSSNSDGTQ 671
 :
 Db 491 EMLHNKCAVAIHQTVFPGACSFDELTKLVROVHHETHHSCSSGSSCSSEANQ 550
 Oy 672 ---SPSLER-PLEDESEIEFNGLEDDEPERECAGSGFGADGDQEVNNAI 721
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 551 ISRAQSLDEFDVLSDEDR---KNTIHLDLQRAD-----SCDGPQYELSETI 593

RESULT 2
 S36616
 regulatory protein sir2 - yeast (Kluyveromyces marxianus var. lactis)
 N:Alternate names: silent information regulator 2
 C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
 C:Accession: A56048; S36616
 R:Chen, X.J.; Clark-Walker, G.D.
 Mol. Cell. Biol. 14, 4501-4508, 1994
 A:Title: sir2 mutants of Kluyveromyces lactis are hypersensitive to DNA-targeting drugs.
 A:Reference number: A56048; MUID:94277055
 A:Accession: A56048
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-670 <CH2>
 A:Cross-references: EMBL:X74569; NID:g397349; PIDN:CMA52661.1; PID:g397350
 C:Superfamily: regulatory protein sir2
 C:Keywords: DNA binding; nucleus; transcription regulation

Query Match 16.1%; Score 621; DB 2; Length 670;
 Best local Similarity 29.3%; Pred. No. 1.8e-31;
 Matches 199; Conservative 95; Mismatches 196; Indels 184; Gaps 26;

Oy 32 LKRRRRDPCGIGRSPGEPSAANVAPAAAGCEAASAA-----APALMREAAGAAASAE 84
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 8 LQKRLE-----QESVAGNGLGESGKAKRGDSVFAARSPENEDVDADAD 55
 Oy 85 REAPATAVAGDNGSGLRREPRADDFD--DEGEED----- 121
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 56 VDADADADADAEADQKDLLEETKADELVEYDEYEKEVSNFNGTASDHVITSSNTG 115
 Oy 122 -EAAAAA---AATGYRDNLLTDLGTLNGHSCESDD--DRTS----- 161
 :
 Db 116 STALASSSDNTNGSGNGTGTWATNGTLSDRQYAPQKPEPIKLERRSVSRKYVEPVISK 175
 Oy 162 ---HASS-----SDWTPRPRIQPYTFVQOHL--MTG---TDPRTL----- 194
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 176 EDSLAKRSTYKQFGSARFLDYLPEDLNSLYY---HMKIKLGFOIKDELMALAOEVYH 232
 Oy 195 ---KDLPE-----TIPPELDMTLMOIVINILSEPPK-----RK 227
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 233 NADNDLSLPQKNSSETKNVSDFYATYPSRPEDPLEKKHAVALKDKQKANKKYLRI 292
 Oy 228 KRKDNITIEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDLPPQA 287
 :
 Db 293 RLJNHTTIDDFPAKLTAKKIIVLGTAGISTSLGIPDRSSGEGFYSKIG-DL-GLNDPQD 350
 Oy 288 MEDIEYFKRDPPEFKFAKEIYPGOFPSLCHKFTALSDKEKLLRNTONTDTEOVAG 347
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 351 VFSLEVFEDBSVFYNIAMVLPENMYSPLSFTKMTQDKKLLRNTQNDINLESTAG 410
 Oy 348 I--QRILQHGSEFATASCLICKYKVDCEAVRGDIFNOVYPRCPRC-----PADP 395
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 411 VEPKAVOCHGSEFATASCTYCHMKIQGERIFPNINLQPLCPYCYSRKLEFFKRTDEE 470
 Oy 396 LA-----IMKEIYFGENLPEQFHRAKMDKDEVDLLIYIGSSL 435
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 471 LADGEDDDMDHGRSVKSPGVLPKPDITTFGEALPSKFHRLIRVDLQCDLLICIGTSL 530

Oy 436 KVRPALLPSSIPHEVQILINREPPLHLPDVELLGDCCVILINELCHRLGE-----YA 490
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 531 KAPVSEIYNMIPNAVPOVLINKDVKHAEPLSLGLCCDDVAAVLAQKCGNDIHDNMN 590
 Oy 491 KLCNPNV-----KLSEITEKPPRPQ--KELVH-----SELPPPLHISED--- 529
 :
 Db 591 KL-KKKVFDESEVERGVYVHPLNESPALLEAEKHLPLQOSTAALRP-PVLSLADSPG 648
 Oy 530 ---SSSPETVPQD 540
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 649 RSSSSSPQPPQTQD 662

RESULT 3
 T39571
 probable regulatory protein sir2-like - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T39571
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.
 submitted to the EMBL Data Library, March 1999
 A:Reference number: 221864
 A:Accession: T39571
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-471 <WOO>
 A:Cross-references: EMBL:AL035637; PIDN:CB38511.2; GSPDB:GN00067; SPDB:SPBC16D10.07c
 A:Experimental source: strain 972h; cosmid cl6D10
 C:Genetics:
 A:Gene: SPDB:SPBC16D10.07c
 A:Map position: 2
 A:Introns: 53/1; 79/1; 106/3; 117/2; 263/1

Query Match 15.7%; Score 605.5; DB 2; Length 471;
 Best local Similarity 34.7%; Pred. No. 1e-30;
 Matches 170; Conservative 70; Mismatches 143; Indels 107; Gaps 19;

Oy 50 PSAANAAPAAAGCEAASAAAPALMREAAGAAASAREAPATVAGDNGSGLRREPRAA 109
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 11 PVASVSPSSG---SSGSALLVDIMCGSKETEDEV----- 44
 Oy 110 DDFDDDEGEEDAEAAAAAAAIYGRDNLLTDLGTLNGHSCESDDDDRTSHASSDWT 169
 :
 Db 45 ---DSDMDKP-----VNDSELENI-----SDLER-----SENV 72
 Oy 170 PPRRIQPYT-FVQOHLMTGTP-RILNDL--LPETIPPELDMTLMOIVINILS-EP 223
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 73 RYLRASGAKFLEKYLIEBELPVRSLKLGINLSAL--EEFEDIDLLPLKLEVLKREV 130
 Oy 224 PKRRKRDINTIEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDL 283
 :
 Db 131 ARIKLPHNTEFEDVNVLLKKAKNVVVLVAGISTSLILFRSDNGFYARLARH--GLS 188
 Oy 284 DQQAQFEDIEYFKRDPPEFKFAKEIYP--GOFOPSLCHKFTALSDKEKLLRNTQNDT 341
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 189 EPSEMFDTHTRENPEITTYTAKDLPLPETNMHSPS--HAFIKLKKKMLSLFTQNDN 246
 Oy 342 LEQVAGI--QRILQHGSEFATASCLICKYKVDCEAVRGDIFNOVYPRCPRC----- 390
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 247 LEKKTGLSDNKILQCHGSEFATATCIKCHKHVDGSLYEDIRNQRVSYCNCKGKPLKLR 306
 Oy 391 -----PADEPLA---IMKEIYFGENLPEQ--FHRAMKDKDEVLLI 429
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 307 VGQNKREKHYSDDGSESEDDLAQPGIMKPDITTFGEALPDSFNKVGSGSELETTDLI 366
 Oy 430 VIGSSLKRVVALIPSSIPHEVQILINREPPLHLPDVELLGD--CDVILINELCHRLG-- 486
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 367 CIGISLKAAPVSELISVIPPTTQIYISRTYVRHQFDVNFSLPCDWIYVEICKRAGWL 426
 Oy 487 GEYAKLCNP 496
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Db 237 RVLNRRVGGLLFRFPDDLDITHDVAKEGSSSSQSSPSASARREGGTEGSSS 296
 QY 466 -----FDVELLDCCVIINELCHRLG--GEYAKLC 493
 Db 297 PNEVEDASTSSSDGYGYGYADYVADYVCRDFFRDCQENYIKLAECLGREALAKRC 356
 QY 494 CNPVKLSITEKPPRPOKELVHSELPT 522
 Db 357 ASPVR-HQLNRHRRMRRE-----SELPPT 380

RESULT 9

S59678
 HS2 protein - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein LPA2c; protein YPL015c
 C:Species: Saccharomyces cerevisiae
 C>Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
 C:Accession: S59678
 R:Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; We submitted to the EMBL Data Library, August 1995
 A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
 A:Reference number: S59677
 A:Accession: S59678
 A:Molecule type: DNA
 A:Residues: 1-357 <HAL>
 A:Cross-references: EMBL:U33335; NID:g965076; PIDN:AB68090.1; PID:g965078; MIPS:YPL015c
 C:Genetics:
 A:Gene: SGD:HS2
 A:Cross-references: SGD:S0005936; MIPS:YPL015c
 A:Map position: 16L
 C:Keywords: transmembrane protein
 F:218-234/Domain: transmembrane #status predicted <TM>

Query Match 11.4%; Score 439; DB 2; Length 357;
 Best Local Similarity 32.7%; Pred. No. 2.1e-20;
 Matches 117; Conservative 58; Mismatches 119; Indels 64; Gaps 11;

QY 247 KIIVLGAGVSVSCGIPDRS-RDGIYARLAVDPDLPDQAMFIEYFRKPRPFKFA 305
 Db 26 KIVFMVAGISTGCGIPDRSPGGLYNHLA--RLKLPYPAVDPVDFQGDPLPYTLA 83
 QY 306 KEIYRGQFOSLCHKFTALSDKEGKLLRNTQNTIDLEOVAGIQ--RLDCHGSFATASC 363
 Db 84 KEIYRGNFRSKEFYLLKFLQDKVLRKYQNTIDLERQAGVADDLIEAHGSFAHCHC 143
 QY 364 LICKYKVCDAVRGDIENQVVRPCRPADPRALIMKPEIYFEGENLPEOPHRAKMYDKD 423
 Db 144 IGCCKVYRPOVYFKSLAEHPIKDFVK--DYCGELVYKALYFEGEDLPDSSETWLNDS 201
 QY 424 -----EVDLLIVGSSSLKVRVALIPSSIPHEVQIILNKEPLPHLHFDVEL 470
 Db 202 WLREKITTSKHNPOQLVIVGTSLAVYRPAISPEIRPKVKRYLCN-----LET 251
 QY 471 LGD-----CDVIINELCHRLGGEYAKLCNPKVLSITEKPPRPQ-----KE 512
 Db 252 VGEFRANKRPDLIV---HQYSDFAEQVLEELGQWDEFEKILTAGGMDNSKEQOLE 307
 QY 513 LVHSELPTPLIATSESSSPERTVPODSSVIATLVQATNNVNDLEVSSEGVEEK 570
 Db 308 IVHDL-----NLSLDQSEHESADKKDKL-----QLNGHSDDEDGASNSSSQK 353

RESULT 10

H69827
 conserved hypothetical protein yhz2 - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: H69827
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertier, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 iech, J.; Harwood, C.R.; Hensel, A.; Hildebrand, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koeltter, P.; Koningsstein, G.; Krogh, S.; Kubano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogihara, B.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schreiber, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchida
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: H69827
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-247 <KUN>
 A:Cross-references: GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAB12804.1; PID:g26333
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yhz2
 C:Superfamily: conserved hypothetical protein b1120

Query Match 9.1%; Score 349; DB 2; Length 247;
 Best Local Similarity 32.7%; Pred. No. 5.7e-15;
 Matches 87; Conservative 54; Mismatches 91; Indels 34; Gaps 9;

QY 235 IEDAVKLLDECKKITIVTACVSVSCGIPDRS-RDGIYARLAVDPDLPDQAMFIEYF 294
 Db 1 METFKSLIHENQRIIVLTGAGMSTESGIPDRSAGIWT-----DASREAM-SLDYF 53
 QY 295 RKDRPPEFKKELY----PGFOPSLCHKFTALSDKEGKLLRNTQNTIDLEOVAGIQ 350
 Db 54 LSYRLFPWKKEKELFOKMSGSEFENEGHLLALEKQKQVDIFTONIDLHKAGSRH 113
 QY 351 ILQCHGSFATASCILCKYKVCDAVRGDIENQVVRPCRPADPRALIMKPEIYFEGENLPEOPHRAKMYDKD 423
 Db 114 VYELHGSTQTNACPAACGARDLP---HLEREVEECTAACNNDDIGTIVKTQVLFQGD 169
 QY 409 NLPRQFRAMKYDR-DEVLDLIVGSSSLKVRVALIP--SSIPHEVQIILNKEPLPHLH 465
 Db 170 AV---MHEDTLYEKLDQADLLVIGTSLVAPARFVPEADALIGMKVITINLEPT----- 222
 QY 466 FDVELLDCCVIINELCHRLGGEYAK 491
 Db 223 -----YCDSLFDMVTHOKIGEFAR 241

RESULT 11

A72370
 regulatory protein, SIR2 family - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: A72370
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72200; MUID:99287316
 A:Accession: A72370
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-246 <ARN>
 A:Cross-references: GB:AE001726; GB:AE000512; NID:g4980992; PIDN:AND35575.1; PID:g498
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TW0490
 C:Superfamily: conserved hypothetical protein b1120

Query Match 9.0%; Score 346; DB 2; Length 246;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:43:37 ; Search time 42.08 Seconds
(without alignments)
565.607 Million cell updates/sec

Title: US-09-461-580A-1

Sequence: 1 MADEVALALQAAAGSPSAAAA.....NEAIRTRQELIDVNPSPDKS 737

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	621	16.1	670	SIR2_KLULA	P33294 Kluyveromyc
2	592	15.4	503	HST1_YEAST	P53665 Kluyveromyc
3	568.5	14.8	515	SIR2_CANAL	O59923 candida alb
4	568	14.7	562	SIR2_YEAST	P06700 saccharomyc
5	484	12.6	381	SIR2_LETMA	Q25337 leishmania
6	439	11.4	357	HST2_YEAST	P53666 saccharomyc
7	341.5	8.9	447	HST3_YEAST	P53667 saccharomyc
8	243.5	6.3	370	HST4_YEAST	P53668 saccharomyc
9	150.5	3.9	3421	TEGU_HSYEB	P28955 equine heip
10	135	3.5	1805	NEST_RAT	P21283 rattus norv
11	133	3.5	1906	KMLS_CHICK	P11789 gallus gall
12	130	3.4	3828	TRX_DROVI	Q24742 drosophila
13	128.5	3.3	1603	PSC_DROME	P35820 drosophila
14	124.5	3.2	3969	HRX_HUMAN	O03164 homo sapien
15	124	3.2	1487	ICP4_HSYEB	P28925 equine heip
16	124	3.2	1487	ICP4_HSYEB	P17473 equine heip
17	121	3.1	627	SPD2_NEPL	P46804 nephila cla
18	119	3.1	1184	FBL2_HUMAN	P58095 homo sapien
19	118	3.1	450	CYL_PARDE	P13627 parascoccu
20	118	3.1	627	YK09_CAEEL	P24304 caenorhabd
21	117.5	3.0	739	BAC1_MOUSE	P97302 mus musculu
22	116.5	3.0	482	YSR2_CAEEL	Q09950 caenorhabd
23	116.5	3.0	606	HMLD_DROAN	P22344 drosophila
24	116	3.0	814	IF39_HUMAN	P55884 homo sapien
25	113.5	2.9	386	HXAD_MOUSE	O62424 mus musculu
26	113.5	2.9	388	HXAD_HUMAN	P12121 homo sapien
27	113	2.9	705	ICAL_BOVIN	P20811 bos taurus
28	112.5	2.9	671	CHS5_YEAST	Q12114 saccharomyc
29	112.5	2.9	1676	APSA_EHENT	Q00083 emericella
30	112	2.9	817	YG4A_YEAST	P46949 saccharomyc
31	111.5	2.9	331	MACS_BOVIN	P12624 bos taurus
32	111.5	2.9	676	ICP0_HSYBK	P29836 bovine heip
33	111.5	2.9	1616	P200_MYCGE	Q49429 mycoplasma

ALIGNMENTS

RESULT	ID	STANDARD	PRT	670 AA.
1	SIR2_KLULA			
AC	P33294			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	REGULATOR PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).			
GN	SIR2.			
OS	Kluyveromyces lactis (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;			
OC	Saccharomycetaceae; Kluyveromyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CBS 2359/152;			
RX	MEDLINE=94277055; PubMed=8007956;			
RA	Chen X.-J., Clark-Walker D.G.;			
RT	"SIR2 mutants of Kluyveromyces lactis are hypersensitive to DNA-			
RT	targeting drugs.";			
RL	Mol. Cell. Biol. 14:4501-4508(1994).			
CC	FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM			
CC	MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS			
CC	INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI			
CC	SILENCING (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: X74569; CAAS2661.1; -			
DR	PIR: S36516; S36516.			
KW	Transcription regulation; Repressor; DNA-binding; Zinc-finger;			
KM	Nuclear protein.			
FT	DOMAIN 44 94 ASP/GLU-RICH (ACIDIC).			
FT	DOMAIN 310 554 SIR2-TYPE CORE DOMAIN.			
FT	ZN_FING 428 455 C4-TYPE (POTENTIAL).			
FT	DOMAIN 650 654 POLY-SER.			
SQ	SEQUENCE 670 AA; 74227 MW; 00A063C/43308B75 CRC64;			
Query Match 16.1%; Score 621; DB 1; Length 670;				
Best Local Similarity 29.5%; Pred. No. 2.1e-30;				
Matches 199; Conservative 95; Mismatches 196; Indels 184; Gaps 26;				
OY	32 LRRPRRDGRLGSPGPSAAVAPAAAGCEAASAA-----APALIMPEAAGAAASAE 84			
DB	8 LQKRPLE-----QESVAENGNGESGKAKGDSVFAARSPENDVDVADAD 55			
OY	85 REAATATVAGDGDGSGRLRPRRAADFD--DEGEED----- 121			

```

DB 56 VADADADADAEEDAKDILEETKADELDEVYDEEKEKVSNNFNCTASDHVGTSSNTG 115
OY 122 -BAAAAA---AALGYRDNLLITDGLTNCFHSCESDD---DRTS----- 161
RT 116 STALASSADTSSGNGCGTAMNTGSDRQAPQKPHPIKLEKRSKRYVFPVISK 175
OY 162 ---HASS-----SDMTPRPRICPPYFVQOHL-MIG---TDPRTL----- 194
DB 176 EDSLNAKRSYLKQFGSARFLDYLPELNSLYV---HMKLGLFOIKDELMALAOEVH 232
OY 195 ---KDLPE-----TIPPELDMTLMQIVINISEPK-----RK 227
DB 233 MADNDSDLPQKASSETKNSDITYATYPSPEDEPLEKHAVALIKDLOKAMNKVLSTRI 292
OY 228 KRKDINTIEDAVKLLQECKKIIVLTGAGVSVCIGIPDFRSRDGIYARLAVDPDLPDQA 287
DB 293 RLINHTTIDDFAKLKTAKKIIVLTGAGISTSLGIPDFRSSGEGYSKIG-DL-GLNDPQD 350
OY 288 MDIEYFKRDPFRFKAKETIPGQFQPSLCKFTALSKKELRLNRYQNTIDTLEQVAG 347
DB 351 VSLVEFTEDEPVSFYNIAMVLPPEMWSPLHSFTKIMIDKRLRNNTQNDINLESYAG 410
OY 348 I--ORILQCHGSFATASCLCKYKVDCEAVRGDIFNQVPRCPRC-----PADPR 395
DB 411 VEPEKMWCHGSFATASCTCHMKIOGERIPFNIRNLQIPCYCKRLEFFKTKTDEE 470
OY 396 LA-----IMKPEIVFGENLPQOFHRAKMYDKDEVDLLIVIGSSL 435
DB 471 LADGEDDDMDHGRHSVPKSFVGLKPDITFFGEALPSKRLHRLREDVLCDDLLICIGTSL 530
OY 436 KVRPAVALPSSIPHEVPOQLINREPLPHLHFDVELLDGCDVIINELCHRLGEB-----YA 490
DB 531 KAAVSEIYNIAPHVQVLINKDPVKAHAEFDLSLGLCDDVAAVAKCGMDIPHDNMN 590
OY 491 KLCGPNV-----KLSEITEKPPRPQ--KELVH-----SELPTPLHISED--- 529
DB 591 KL-KKKVDPSEVEKGVKVPKPLNSPALLEAEKEKHLPLOOSTALTP-PVLSLSDPG 648
OY 530 ---SSPERTVPOD 540
DB 649 RSSSSPPPTQTD 662

```

```

RX MEDLINE-97321807; PubMed-9178509;
RA Tzeremia M., Katsoulou C., Alexandraki D.;
RT "Sequence analysis of a 3.2 kb segment from the left arm of yeast
RT chromosome XV reveals eight known genes and ten new open reading
RT frames including homologues of ABC transporters, inositol
RT phosphatases and human expressed sequence tags.";
RL Yeast 13:583-589(1997).
CC -I- FUNCTION: INVOLVED IN TELOMERIC SILENCING, AS OVEREXPRESSION
CC RESTORES SILENCING AT HMR IN SIR2 MUTANTS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -I- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U39041; AAA81033.1; -
DR EMBL; L47120; AAB38430.1; -
DR EMBL; Z74810; CAB9078.1; -
DR SGD; S0005429; HST1.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 200 439 SIR2-TYPE CORE DOMAIN.
FT ZN-FING 318 345 C4-TYPE (POTENTIAL).
SQ SEQUENCE 503 AA; 57702 MW; 4CDF2799E4135ABB CAC64;

```

Query Match 15.4%; Score 592; DB 1; Length 503;
Best Local Similarity 37.3%; Pred. No. 8.1e-29;
Matches 138; Conservative 62; Mismatches 104; Indels 66; Gaps 9;

```

OY 188 TDPR-----TLKDLPEPTIPPELDMTLMQIYIN-ILSEPPKRRKRDINTEDAVK 240
DB 155 TDPLEKHAVALIKDL-----QKAINKVL---TRLRFPNENTIDHFTA 195
OY 241 LLOECKKIIVLTGAGVSVCIGIPDFRSRDGIYARLAVDPDLPDQAMPDIYFPRDPR 300
DB 196 TLRNKKIIVLTGAGVSTSLGIPDFRSSGEGYSK--IRHLGLEDDVDVFNLDIFLDPSV 253
OY 301 PFKFAKEIYPOGQFQPSLCKFTALSDKEGKRLNRYQNTIDTLEQVAGI--QRILQCHGSF 358
DB 254 FYNIAMVLPPEMWSPLHSFTKIMLODKGKRLRNNTQNDINLESYAGIDPKLVQCHGSF 313
OY 339 ATASCVTKHMOIPGKRIEININLEPLCPCYCKRKQYFPMSNGNNTVQTNINFNSTPL 373
DB 314 ATASCVTKHMOIPGKRIEININLEPLCPCYCKRKQYFPMSNGNNTVQTNINFNSTPL 373
OY 394 EPLALMKPEIVFGENLPQOFHRAKMYDKDEVDLLIVIGSSLKVRPAVALPSSIPHEVPO 453
DB 374 KSYGVAKPDMPTFFGALSRFHKTIRKDLBCDDLICIGTSLKAVPVSITYAMVSHVPQ 433
OY 454 ILINREPLPHLHFDVELLDGCDVIIN-----ELCHRLGGEYAKLCCNPYKLSERT-- 503
DB 434 ILINRDVYTHAHEFDLNLGFCODVASIAKKCHMWIPIPKKWDLKKIYDNCDEIDKGYTK 493
OY 504 -EKPPRPOKE 512
DB 494 IKKQPRKKQO 503

```

```

RESULT 2
HST1_YEAST STANDARD: PRT; 503 AA.
AC P33685;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST1 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 1).
GN HST1 OR YOL066C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / YPH1;
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GRY 668;
RX MEDLINE-96405912; PubMed-8810037;
RA Derbyshire M.K., Weinstock K.G., Strathern J.N.;
RT "HST1, a new member of the Sir2 family of genes.";
RL Yeast 12:631-640(1996).
RN [3]
RP SEQUENCE FROM N.A.

```

OS Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC anamorphic Saccharomycetales; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SC5314;
 RA Perez-Martin J., Johnson A.D.;
 RT "SIR2 gene from Candida albicans";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF045774; AAC09304.1; -
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN.FING 352 379 C4-TYPE (POTENTIAL).
 FT 2N.FING 352 379
 SQ SEQUENCE 515 AA; 57732 MW; 2E4DB259EFA9251F CRC64;

Query Match 14.8%; Score 568.5; DB 1; Length 515;
 Best Local Similarity 39.4%; Pred. No. 2.2e-27;
 Matches 141; Conservative 60; Mismatches 122; Indels 35; Gaps 10;

QY 138 LLLTDLGTLNHFHSCSD--DRTSHA---SSSDWTPRRPRTIGTYTQOHLMTGTPRT 192
 DB 141 LMLTDSQLEPKYEIDNDKGTNNSDIDSNDMSQSESG-----ELGDAMD-- 190
 QY 193 ILKDLPEITPPPELD-DMTLMQIVINILSEPPRRKKRKINTEDAVKLQECCKIIVL 251
 DB 191 -VDSLSPEEN--EDDYDDDMSTTLKRTINMPFKYK-----LPDLISDSRAKKIMVY 240
 QY 252 TGAGVSVSCGIPDFRSRDGIYARLAVDPDLDPQAMFDIEYFRKDPPEFFKAKEIYPG 311
 DB 241 TGAGISTSLGIPDRSPFKGLYNQLSK--LNLSDPQKVFQDLQTFMRGRLFTIAHLVLP 298
 QY 312 QFQPSLCHKFIALSDKGLKRLNTONTIDLEOVAGI--QRILOCHGSFATASCLICKYK 369
 DB 299 DGRKSLHAFLKLLQDNHKLRLNTONTIDLEOVAGI--QRILOCHGSFATASCLICKYK 358
 QY 370 VDEAVAGDIFNQVPRPCRC-----PADEPLAIMEKEIYFEGNLEPQOFRAKKYDDE 424
 DB 359 FAGKIKTNHRRKQVPCALCMKTKQAPRIHFGAIKPTITTFEGEDLPERFTITLMDKDLQ 418
 QY 425 VDLIIVIGSSILKRVPAVLIPISSIPHEVPQIILNREPLPHLHFDVLLGDCDVIINELC 482
 DB 419 IDLFVLIGTSLKVPVASSIIEYRVKPKILINDPIPNRGFNQLGLGCDAAVSYLC 476

RESULT 4
 SIR2_YEAST
 ID SIR2_YEAST STANDARD; PRT; 562 AA.
 AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 OS SIR2 OR MARI OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85126876; PubMed-6096447;
 RA Shore D., Squire M., Nasmyth K.A.;
 RT "Characterization of two genes required for the position-effect
 RT control of yeast mating-type genes";
 RL EMBO J. 3:2817-2823(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Paulin L., Saren A.M., Laamanen P.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES. IT IS
 CC INVOLVED IN TELOMERIC SILENCING AND IN HM MATING TYPE LOCI
 CC SILENCING. SUPPRESSES RNA RECOMBINATION AND PROMOTES HISTONE
 CC DEACETYLATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X01419; CAA25667.1; -
 DR EMBL: Z71781; CAA96447.1; -
 DR EMBL: Z74090; CAA98600.1; -
 DR PTR: S05891; RGSY52.
 DR SGD: S0002200; SIR2.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN.FING 372 399 C4-TYPE (POTENTIAL).
 FT 2N.FING 372 399
 SQ SEQUENCE 562 AA; 63261 MW; 52E693753654586 CRC64;

Query Match 14.7%; Score 568; DB 1; Length 562;
 Best Local Similarity 39.1%; Pred. No. 2.6e-27;
 Matches 127; Conservative 46; Mismatches 92; Indels 58; Gaps 7;

QY 216 VINILSEPPKKRRK---KDIN-----TIEDAVKLLQECCKIIVLTG 253
 DB 205 VTNV--EDPLAKQGVRLIKDQRAINKVLTCTRLRLSNFTIHDHFKLHTARKILVLTG 262
 QY 254 AGVSVSCGIPDFRSRDGIYARLAVDPDLDPQAMFDIEYFRKDPPEFFKAKEIYPGQF 313
 DB 263 AGVSTSLGIPDRSPFKGLYNQLSK--LKHGLDLPQOVFNINIMHPSPVFNANNVLP 320
 QY 314 QPSLCHKFIALSDKGLKRLNTONTIDLEOVAGI--QRILOCHGSFATASCLICKYKVD 371
 DB 321 IYSLHSTIKMLQMKGLKRLNTONTIDLEOVAGI--QRILOCHGSFATASCLICKYK 380
 QY 372 CEAVAGDIFNQVPRPCRC-----PADEP-----LAIKRP 401
 DB 381 GERIFNKRIRNLEPLCPYCKYKRRREYFEGYNNKVGVAASGMSSERPYILNSGYLKP 440
 QY 402 EIVFEGNLPQOFRAKKYDDEVDLIVIGSSILKRVPAVLIPISSIPHEVPQIILNREPL 461
 DB 441 DITFEGEALPKRFRKSIDIEDLCEDLICIGTSLKVPVASSIIEYRVKPKILINDPIPN 500
 QY 462 PHLHFDVLLGDCDVIINELCHRLG 486
 DB 501 KHAEPDLSLGYCDIAMAAYAKCG 525

RESULT 5
 SIR2_LEIMA
 ID SIR2_LEIMA STANDARD; PRT; 381 AA.
 AC P06700;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 (SILENT INFORMATION REGULATOR 2).
 OS SIR2 OR MARI OR YDL042C OR D2714.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

AC Q25337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE REGULATORY PROTEIN SIR2 HOMOLOG (LMSIR2RP).
 GN SIR2.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 RN [1]
 RP MEDLINE=96186914; PubMed=8635734;
 RA Yahiaoui B., Taibi A., Ouassil A.;
 RT "A Leishmania major protein with extensive homology to silent
 RL Gene 169.115-118(1996)."
 CC -1- FUNCTION: KEY REGULATOR OF GENE EXPRESSION BELIEVED TO HELP FORM
 CC MODIFIED CHROMATIN STRUCTURES ON THE GENES IT REGULATES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, L40331; AAB06804.1; -
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 FT DOMAIN 31 243 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 148 151 POLY-ALA.
 FT DOMAIN 152 179 C4-TYPE (POTENTIAL).
 FT DOMAIN 271 278 POLY-SER.
 FT DOMAIN 305 310 POLY-SER.
 SQ SEQUENCE 381 AA; 41958 MW; 5A311630A5D2365 CRC64;

Query Match 12.6%; Score 484; DB 1; Length 381;
 Best Local Similarity 32.1%; Pred. No. 1.9e-22;
 Matches 125; Conservative 55; Mismatches 101; Indels 108; Gaps 12;

DB 220 LSEPKRRKKRDIETEDVAKLLE--CKIIVLFGAGVSGGIPDFERSD-GIYARLA 276
 14 LGEP-----IVEGLARIKEDVARRILVVGAGSVAGIPDFERSDGIYAKLG 63
 QY 277 VDEPDLDPQAMFIEYFRKDRPFEPFAKE--IYPGOFQPSLCHKETALSDEKGLRN 334
 DB 64 K--VYNLDPTDAFSLTLLEKEPEIFYSARELMLMGPHQPTAVHHFILLQDEGRLLRC 121
 QY 335 YTONIDTLEQVAGI--QRLQCHGSPATASCLICKYKDCVCEAVRGDIFNOVYPRCPRA 392
 DB 122 CQNQNDIGLEKAGVSPPELLVEAHGSAACIECHTFSEIDNYLEAMSGTVSRCTCG- 180
 QY 393 DEPLAIKPEIYFGEFNLPEOFHRAKKYDKDEVLLIVGSSILKRVPLVLISSIPHEVP 452
 DB 181 ----GIVKKNVYFGEENLPDAFPAALHDAPIAEIVITIGTMQVHPRFALLCVPKSP 236
 QY 453 QILNRE-----PLPHLH----- 465
 DB 237 RVLNMRERVGGLLFRPDDPLDTHDDAVAKEGRSSSSQSRPSASAREEGTDEGSSS 296
 QY 466 -----PDVELLGDODVLIINELCHNLG--GEYAKIC 493
 DB 297 PNEEVEDASTSSSDGYGYGYVYAHPDVCRDVFYFRGDCQENVLKLAECLGREALAKRC 356
 QY 494 CNPKLSITEKPRPRQKELVHLSLPT 522
 DB 357 ASPVR-HQLHRRRRRRR-----SELPT 380

RESULT 6
 HST2_YEAST
 ID HST2_YEAST STANDARD; PRT; 357 AA.
 AC P53686;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HST2 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 2).
 GN HST2 OR YPL015C OR LPAC2.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / YPH1;
 RX MEDLINE=96101589; PubMed=7498786;
 RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
 RA Boeke J.D.;
 RT "The SIR2 gene family, conserved from bacteria to humans, functions
 RT in silencing, cell cycle progression, and chromosome stability.";
 RL Genes Dev. 9:2888-2902(1995).
 CC [2]
 CC SEQUENCE FROM N.A.
 RA Hall J., Ahmed A., Bussey H., Fortin N., Friesen J.D., Storms R.K.,
 RA Vo D.H., Wang Y., Winnett E.;
 RL Submitted (Aug1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD FUNCTION IN TELOMERIC SILENCING, CELL CYCLE
 CC PROGRESSION AND CHROMOSOME STABILITY.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, U39063; AAB81035.1; -
 DR EMBL, U33335; AAB68090.1; -
 DR SGD: S0005936; HST2.
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT DOMAIN 22 250 SIR2-TYPE CORE DOMAIN.
 FT ZN_FING 143 173 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 357 AA; 39979 MW; ED281E5B8241A4D0 CRC64;

Query Match 11.4%; Score 439; DB 1; Length 357;
 Best Local Similarity 32.7%; Pred. No. 9.1e-20;
 Matches 117; Conservative 58; Mismatches 119; Indels 64; Gaps 11;

DB 247 KIIVLFGAGVSGGIPDFERS-RDGITARLAVFDPDLDPQAMFIEYFRKDRPFEPFA 305
 26 KVIYFAGAGISTSCGIPDFERSPTGLYNLA--RLKLPYEAIVDFQSDPLPEYTLA 83
 QY 306 KEIYGGQPSLCHKETALSDEKGLRNQYTONIDTLEQVAGI--QRLQCHGSPATASC 363
 DB 84 KELYFGNEFSPKFHYLKLKEQDKVAKRYTONIDTLEQAGVAKDDLIIEAHGSAFHCHC 143
 QY 364 LICKYKVCCEAVRGDIFNOVYPRCPADPLAIKPEIYFGEFNLPEOFHRAKKYDKD 423
 DB 144 ICGKAVYPPQVFKSLAEHRPIDFYKC--DVGELVKAIVAFEGEDLPDSFETMLNSE 201
 QY 424 -----EVDLLIVGSSILKRVPLVLISSIPHEVPQILINREPLPHLFDVEL 470
 DB 202 WLREKITTSKGRPOQPLVIVGTSIAVYFPAFLPEIDPRKAVRYLCN-----LET 251
 QY 471 LGD-----CVIINELCHNLGGEYAKICNPVLSITEKPRPRQ-----KE 512
 DB 252 VGDFRANKRPDILV----HQSDEFAQOLVEELQMOEDFEKILTAQSGMGDSKEOLLE 307


```

OY 513 LVHLSLPPPLHISSESSPERTVPODSSVATLVDOATNNVNDLEVSSESCVEEK 570
DB 308 IVDLE-----NLSLDQSEHSADKKDKL-----QRLNGHSDSDGANSNSSSK 353

RESULT 7
HST3_YEAST STANDARD: PRT: 447 AA.
ID HST3_YEAST
AC P53687:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST3 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 3).
GN HST3 OR YOR025W OR OR26.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / YPH1;
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RA de Haan M., Maarse A.C., Grievell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST4 TO TELOMERIC SILENCING,
CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
CC AND GENOMIC MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39062; AA81034.1; -
DR EMBL: X87331; CA65074.1; -
DR EMBL: Z74933; CA93215.1; -
DR SGD: S0005551; HST3.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 50 314 SIR2-TYPE CORE DOMAIN.
FT ZN_FING 195 223 C4-TYPE (POTENTIAL).
SQ SEQUENCE 447 AA; 50524 MW; 59CD5FBD712B7005 CRC64;

Query Match 8.98; Score 341.5; DB 1; Length 447;
Best Local Similarity 28.38; Pred. No. 9,4e-14;
Matches 119; Conservative 60; Mismatches 147; Indels 95; Gaps 19;

OY 231 DIMTIEDAVLLOECKKIIVLTGAGVSGIPFRSRDGIYARLAVD---FPDLPPOA 287
DB 38 DDEVLRVTVQLSRKRIACITGAGISCMAGIPFRSSDGLYDKDCSQYMSIKSGRE 97
OY 288 MFDEYERKDP--PEFKAKEIYPG--OFOPSLCHKFIALSDKEGKLRLRYTONIDTLE 343
DB 98 MFDEISLRDFPKISIFAKFMEKRLSVNQVLAQKPTKHKFIALKRNKRLRCYTONIDGLE 157
OY 344 QVAGIQ-----RIIQGHSFATASCLIC-----KIVKDEAVRGD 378
DB 158 EISGILTLNKRKLPLTSPSSHKKNLDVVOYHGDLEKTLSCTKFQTPMRSRYMSRC-LRGE 216
OY 379 IFNQVVRCPRCRPA-----DEPLAIKPEIVFGENLP---EQGHRAMKTD-- 421

```

```

DB 217 L-----PLCPDCEALINKRLNEGKRTIGSNVGLRPNIVLYGENHSPCEITIGLNDII 271
OY 422 KDEVDLIVIGSSLKVRPVALIPSSIPHEVPO-----ILNRELP-----HLHFDVELG 472
DB 272 KGNPDFLIMOTSLKVGVOVQLVKRLSKKIDRGLIILVKNPKPIGSSWGIIIDVOIHS 331
OY 473 DCDVITNELCHRIIGGEYAKLCCNPKLSEI-----TEKPPQKELVHSELPPPLHI 526
DB 332 DCDMMWTFLESQI-PDFKTIQDQIKRLRQKREASDLRKQKAKQDSI---GTPPT---- 383
OY 527 SEDSSPERTVPODSSVATLVDOATNNVNDLEVSSESCVEEK---POEYQTSRNVENI 583
DB 384 -----TLRT-----AGSIDQGNNELN-TIKSLNTYKRIILSPENSEDEEENL 429
OY 584 N 584
DB 430 D 430

RESULT 8
HST4_YEAST STANDARD: PRT: 370 AA.
ID HST4_YEAST
AC P53688;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HST4 PROTEIN (HOMOLOGOUS TO SIR2 PROTEIN 4).
GN HST4 OR YDR191W OR YD9346.03.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP GENE NAME.
RC STRAIN-GRY 668;
RX MEDLINE-96405912; PubMed-8810037;
RA Delvishire M.K., Weinstein K.G., Strathern J.N.;
RT "HST1, a new member of the SIR2 family of genes.";
RL Yeast 12:631-640(1996).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-96101589; PubMed-7498786;
RA Brachmann C.B., Sherman J.M., Devine S.E., Cameron E.E., Pillus L.,
RA Boeke J.D.;
RT "The SIR2 gene family, conserved from bacteria to humans, functions
RT in silencing, cell cycle progression, and chromosome stability.";
RL Genes Dev. 9:2888-2902(1995).
CC -1- FUNCTION: CONTRIBUTES TOGETHER WITH HST3 TO TELOMERIC SILENCING,
CC PROPER CELL CYCLE PROGRESSION, DNA DAMAGE CONTROL, RECOMBINATION,
CC AND GENOMIC MAINTENANCE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SIR2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z48784; CA88705.1; -
DR SGD: S0002599; HST4.
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT DOMAIN 90 342 SIR2-TYPE CORE DOMAIN.
FT ZN_FING 221 254 C4-TYPE (POTENTIAL).
SQ SEQUENCE 370 AA; 41765 MW; 86BB0238BFA914F1 CRC64;

```


RX MEDLINE-90150286; PubMed-1689217;
 RA Lendahl U., Zimmerman L.B., McKay R.D.G.;
 RT "CNS stem cells express a new class of intermediate filament
 RL protein.";
 CC Cell 60:585-595(1990).
 CC -1- TISSUE SPECIFICITY: CNS STEM CELLS.
 CC -1- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
 CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; M34384; AAA41685.1; .
 DR PIR; A34736; A34736.
 DR INTERPRO; IPR001664; .
 DR PRAM; PF000038; filament; 2.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Heptad repeat pattern;
 KM Neurone.
 FT DOMAIN 1 7 HEAD.
 FT DOMAIN 8 314 ROD.
 FT DOMAIN 315 1805 TAIL.
 FT DOMAIN 8 43 COIL 1A.
 FT DOMAIN 44 55 LINKER 1.
 FT DOMAIN 56 151 COIL 1B.
 FT DOMAIN 152 174 LINKER 12.
 FT DOMAIN 175 193 COIL 2A.
 FT DOMAIN 194 196 LINKER 2.
 FT DOMAIN 197 314 COIL 2A.
 SQ SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;
 Query Match 3.58; Score 135; DB 1; Length 1805;
 Best Local Similarity 19.7%; Pred. No. 1.7; Mismatches 262; Indels 272; Gaps 36;
 Matches 155; Conservative 99;

QY 471 LGDCDVIINELCHRLGGEYAKLCNPKVLSITEKPPR-OKELVHLSLPP- 522
 DB 756 ---DQMLERLEKESQSLK---SPENQRIKPLEENKSLRYLEENOTVPLESR 808
 QY 523 -----PLH-----ISDSSSPERTVQDSSVIAITLVDAATNN 554
 DB 809 NQRLRLSLEVEERQRIYKPLEKXSQSLGSLAENNOPLVLEEDCINSLLEDKTHK 868
 QY 555 NVNDLE-----VSESSCYEKQEVQTSRNNENINVENPDKRAVSSADKNER 603
 DB 869 SLGLEDRMNGDSIIIPQSEFTQ-VSLRPEEEDORIVHLEKESQEF-----SR 916
 QY 604 TSVAETKCKPNPLAKQISK-LEGQYLFVPPNRIFFGAEVYSDSEDDVLSSSSG 662
 DB 917 SSEE-----EQVMSRLEGEN-----HESLSVKEEDQWVESQLEK 953
 QY 663 SNSDGTGQSPLEPLEDESEIEFYNGLEDTERPEPCAGSGFGADGQEVVNEAIA 722
 DB 954 ESQDSG-----KSLEDS--QETGPLEKNAE-----SURLAGQDQDEQKLEQ 997
 QY 723 TRQELTDV 730
 DB 998 TQQTLLAV 1005
 RESULT 11
 ID KMLS_CHICK STANDARD; PRT; 1906 AA.
 AC P11799; P19038;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMS
 DE (EC 2.7.1.117) (MCK) [CONTAINS: TELOKIN].
 OS Gallus gallus (chicken).
 OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A. (MCK-210).
 RX MEDLINE-96033976; PubMed-7589469;
 RA Watterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,
 RA Stepanova O.V., Shtrinsky V.P.;
 RT "Multiple gene products are produced from a novel protein kinase
 RT transcription region.";
 RL FEBS Lett. 373:217-220(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (MCK-108).
 RX MEDLINE-90192792; PubMed-2315320;
 RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
 RA Means A.R.;
 RT "Regulatory and structural motifs of chicken gizzard myosin light
 RT chain kinase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
 RN [3]
 RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=EMBROBLAST;
 RX MEDLINE-90361738; PubMed-2202734;
 RA Shemakher M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
 RA Marisjan P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
 RA van Eldik L.J., Watterson D.M.;
 RT "Use of DNA sequence and mutant analyses and antisense
 RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
 RT myosin light chain kinase autoinhibition, calmodulin recognition, and
 RT activity.";
 RL J. Cell Biol. 111:1107-1125(1990).
 RN [4]
 RP SEQUENCE OF 1259-1906 FROM N.A.
 RC TISSUE=GIZZARD;
 RX MEDLINE-87157587; PubMed-3030394;
 RA Guerriero V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;
 RT "Domain organization of chicken gizzard myosin light chain kinase

RT deduced from a cloned cDNA.";
 RL Biochemistry 25:8372-8381(1986).
 [5]
 RN SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RP TISSUE-GIZZARD:
 RC MEDLINE-93073972; PubMed-1444462;
 RA Yoshikaki S., Ikebe M.;
 RT "Molecular cloning of the chicken gizzard telokin gene and cDNA.";
 RL Arch. Biochem. Biophys. 299:242-247(1992).
 [6]
 RN SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RP MEDLINE-92236611; PubMed-1373815;
 RA Collinge W., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
 van Eldik L.J., Waterson D.M.;
 RT "Structure and expression of a calcium-binding protein gene contained
 within a calmodulin-regulated protein kinase gene.";
 RL Mol. Cell. Biol. 12:2359-2371(1992).
 CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A
 MYOSIN LIGHT CHAIN, WHICH LEADS TO THE FORMATION CALMODULIN/
 MYCK SIGNAL TRANSDUCTION COMPLEXES WHICH ALLOW SELECTIVE
 CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN LIGHT-CHAIN] = ADP + [MYOSIN
 LIGHT-CHAIN] PHOSPHATE.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS OF MLCK ARE PRODUCED
 BY ALTERNATIVE INITIATION, A NON-MUSCLE FORM (MLCK-210; WHICH IS
 CC THE LONGEST FORM); A SMOOTH-MUSCLE FORM (MLCK-108) AND TELOKIN (A
 CC C-TERMINAL SECTION WITH NO CATALYTIC ACTIVITY).
 CC -1- TISSUE SPECIFICITY: TELOKIN IS EXPRESSED IN GIZZARD, HEART, LUNG,
 CC INTERESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF THE
 CC EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE GIZZARD.
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 CC -----
 DR EMBL; X52876; CAA37056.1; -;
 DR EMBL; X52876; CAA37057.1; -;
 DR EMBL; X52876; CAA37058.1; -;
 DR EMBL; M31048; AAA49069.1; -;
 DR EMBL; M14953; AAA69964.1; -;
 DR EMBL; M96855; AAA49083.1; -;
 DR EMBL; M88283; AAA48647.1; -;
 DR EMBL; M88284; AAB53768.1; -;
 DR PIR; A25810; A25810.
 DR PIR; A37099; A37099.
 DR PIR; S11652; S11652.
 DR INTERPRO: IPR000719; -;
 DR INTERPRO: IPR001777; -;
 DR INTERPRO: IPR002290; -;
 DR INTERPRO: IPR003006; -;
 DR PFAM: PF00041; fn3; 1.
 DR PFAM: PF00047; Ig; 9.
 DR PFAM: PF00069; pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
 KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 KW Alternative Initiation.
 FT CHAIN 1 1906 MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
 FT CHAIN 935 1906 MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
 FT CHAIN 1750 1906 TELOKIN.

Query Match	Score 133; DB 1; Length 1906;
Best Local Similarity 19.0%; Pred No. 2.5; Mismatches 288; Gaps 35;	
Matches 150; Conserved 95; Mismatches 257; Indels 288; Gaps 35;	
65 SAAPAAIMREAAAGAAASAREAPATAVAGDGN-----GSGILRR---BPRADDPD 113	FOR MLCK-108.
827 SSASRAEMLRDRESASGERR-----DCGNVGFRTSGPKKSSSETRAEEQ 878	FOR TELOKIN.
114 DD-----EGEEDDAAAAAAAAAIGYDNLITGLTNGFHCSEDDDDRTSIA 163	FIBRONECTIN TYPE-III.
879 EDVGVLRKRVETREHTESLRQEAELDFRD--ILGRKYSTKSP---SEEDLKEIPA 932	PROTEIN KINASE.
164 SSDMTP--RPRIGPYTFVQO---HLMIGTDPRTIL-----KDLPEPTIPPELDWT 211	IG-LIKE C2-TYPE DOMAIN.
933 EGMDFRANLQROVKRKTLSSEERKHAAPQVDFRSLAKKGPRTPLPKVPPK-PAVT 991	ATP (BY SIMILARITY).
212 LMQIYINILSEPPKRRKRKDINTIEDAVKLLQCKKIIVLTGAGVSSGCIPIDFRSRDI 271	ATP (BY SIMILARITY).
992 DRSVYLGAKKKPPAE-----NGSASTPAPNARAGS 1021	CALMODULIN AUTOINHIBITION (AM13) REGION
272 VARLAVDFPDLDPQAMPDIEFRKDRPFKAKETIYGCQFQPSLCHFIALSKEGL 331	(POTENTIAL).
1022 EAQNAITPNSAP-----APKPVK-----KEEN 1045	CALMODULIN RECOGNITION (RS20) REGION
332 LKNYQNIDTLBQVAGIQRILOC-HGSFATASCLCKYKVDDEAVRGDIFNOVPRCPRC 390	(POTENTIAL).
1046 DR-----KCEHG-----CAVVDGGITGKAKKPPA 1071	MOTIF IA.
391 PADEPLAKKPEIVFEGENLPEQFHRAMKYDDEVLLIVISLSLKVRAVAILPSSIPHE 450	MOTIF IB.
1072 SKPTP-----PSPKGTAPSFTEKLDKAKVADEKILVLCGRISDPASVSWLDSKAIKS 1126	4 X REPEATS, MOTIF IIA.
451 VQOILNREPDLPHLHEDVLLAD-CDVITINELCHRLGEGY-----AKLCC----- 494	IIA-1.
1127 SKSIYISGE-----GLCSLTIEKVMPEGGEGYKCIANAAGKACACAKVIVE 1174	IIA-2.
	IIA-3.
	IIA-4.
	5 X REPEATS, MOTIF IIB.
	IIB-1.
	IIB-2.
	IIB-3.
	IIB-4.
	IIB-5.
	4 X REPEATS, MOTIF III.
	III-1.
	III-2.
	III-3.
	III-4.
	CALMODULIN-BINDING.
	POLY-GLU.
	PHOSPHORYLATION.
	PHOSPHORYLATION.
	R -> Q (IN REF. 4).
	SEQUENCE 1906 AA; 210445 MW; AD7DBA3B69EE3363 CRC64;

```

OY 495 --NPVLTSETEKPPRQKELVHLSLPTPLHISDSSSPERTVPODSSVATLVQAT 552
DB 1175 DRSSTAAKAEKTKKPK-----TLLPVLSTESSEPAVKKPAFTPPKAAFT--PPQI 1227
OY 553 NNNVNDLEVESSESCVEKEPEOVOTS-----RVENINVENPDKFVAGSSTAD 539
DB 1228 TOPEDRKVRAGSEVLEFAKVGTAITPCWMMFKRQIQENEXIKIEN-----AE 1277
OY 600 KNERISVAETVRK---CW-----PNRLAKEDISRLBGNQYLV---PPNRYTFHGAENV 648
DB 1278 NSSKLTISSTKOEHCCCYTLVENKLGSRQAQVNL-----TVVDKPPPP-----AGTPCA 1327
OY 649 SDESDVVLSSSCGNSDSG-----TCOSPS----- 674
DB 1328 SDRSSSLTLWYSGSSIDGSAVQSTVETWNSVDNKMTDLTTCRSTSFVWQDLQADREY 1387
OY 675 -----LEPPEDESEIEEFYNGLEDPTREPECAGSGFGADGQDEVVN---EA 720
DB 1388 KFRVRANVYGISP--SQSEVYKVGKQEBELKEE---AELSDDEGKETEVNRYTVT 1442
OY 721 IATROELTVD 730
DB 1443 INTEQKVDV 1452

RESULT 12
TRX_DROVI STANDARD; PRT: 3828 AA.
AC 024742;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRITHORAX PROTEIN.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RL between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (Ant-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-FINGER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 250038; CAA90349.1; -
CC DR HSSP: P04002; IMFA.
CC DR FLYBASE: FBgn0014844; Dvir\trx.
CC DR INTERPRO: IPR001214; -
CC DR INTERPRO: IPR001965; -
CC DR PFAM: PF00628; PHD; 2.
CC DR PFAM: PF00856; SET; 1.
CC KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
CC Nuclear protein; Developmental protein; Activator.
CC DOMAIN 1251 1334 PHD 1.

```

```

FT DOMAIN 1335 1380 PHD 2.
FT DOMAIN 1408 1469 PHD 3.
FT DOMAIN 1708 1767 PHD 4 (ATYPICAL).
FT DOMAIN 1768 1818 PHD 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET DOMAIN.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-ALA.
FT DOMAIN 243 251 POLY-GLN.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 538 546 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
SQ SEQUENCE 3828 AA; 413721 MW; 32059CF30A3C504 CRC64;

Query Match 3.4%; Score 130; DB 1; Length 3828;
Best Local Similarity 18.1%; Pred. No. 9.3;
Matches 142; Conservative 103; Mismatches 282; Indels 258; Gaps 36;

OY 20 AMEASQPADEPLKRRRDPGPGIGRSPGERSAANVAPAAACCEAASAAAPALMREAGA 79
DB 469 SVHASSNSIOESSSTTNLFGSGLSRAFGAPIDD-----EDAVSG 508
OY 80 AASAREAPATAVAGDG-DNGSGLRREPRAADPDDDEGEDEEAASAAAAGYDNL 138
DB 509 GVTFRKPEQHKPTEDNDGDSASDAIEDDEDIDDDAEENEAASEKSETTASVDEK 568
OY 139 LITDGLITNGFHSCSDDDRTSH-----ASSDWTTPP----- 172
DB 569 EADDROQLVMKHFLVLPKSTRSSRIIPKNRRLLEVGGICSKRSPSDANGKPKPNYFGLA 628
OY 173 -----RIGYTFV-----QOHMIGTD-PTIL-K 195
DB 629 TLPAKCTPRRRRSANATLSKLGKETPASATAKVNSFVLRQRFQFDKSSFPVSAK 688
OY 196 DLLP-ETIPPELDMTLMOIVINILSEPPKRRKRDINTEDAVKLLQECKIIVLTGA 254
DB 689 PLEPTTVTLAASSAII-----SANVLFGA-----LNANSAVMAASTC----- 728
OY 255 GVSVSGCIPDFRSHDGIYARLAVDPDLPDQAMEDIETFRKDPPEFKAKETIPQOFQ 314
DB 729 ---AVCSAP-VNNKADAPLAR-----KY-----GVIA 750
OY 315 PSICHRFIASDKEGKLLRNTQNDIDLEQVAGIQRILQCHGSFATASCILCKKYVCEA 374
DB 751 CEVCRKFNSSMTKISKISTPMHNSPSTSTASGQO-LKCTDG---GNCSTLSLSQLN 805
OY 375 VRGDFNQVVRPCRPCPADEPLAIKPEIYFEGENLPEGFHRAKVKDDEVLLIVIGSS 434
DB 806 FK-KLYKE---KCKACKLKKCLATLQ-----LAG-HRS----- 834
OY 435 LKVRPAVALIPSSIPHEV-----PQILINREPLPHLF-----DVELLGCDV 476
DB 835 ---RLSAIPLASMEVEAVAPDDKCPPELL---SPFASIRFTAPTSSASGGTIKKWSAET 888
OY 477 IINEL-CHRLGGEYAKLCCPNVAKISETTEKP-----PRQKELVHLSLPTPL 524
DB 889 AVNSIKSNPLAENNVTGCTPLRALPLEKPLFLKIGSDNKKAKESKALGLSPVST- 946
OY 525 HISDSSSPERT---VPQDSSVATLVQATNNVNDLEVESSESCV-----EKPOEVOTS 577
DB 947 --SEAAVAPGKTRKKAQODEK-ARELEAKPLSPNAKKTTTEANTPTQODEOPASTTTT 1003
OY 578 RVENINVENPDKFVAGSSTADKNERTSVAETVRKCPNRLAKEDISRLBGNQYLVFP 637
DB 1004 VSAASSTSHTSAAATNSQLETTAAANASAV-----PDNLKRRRID--LKG-----PR 1050
OY 638 NRTIFHGAEV-----YDSEDDVLSSSSCGNSDGTGQSPSL-----EPLP- 680

```

```

Db 1051 VKHVRSAISIVLGOPLATFGDEEELAAAEAPAPTTTTTTTSPVITIKKPKSPQPMOMI 1110
Oy 681 -DESE 684
Db 1111 IDEND 1115

RESULT 13
PSC_DROME STANDARD: PRT: 1603 AA.
AC P35820.1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-1994 (Rel. 29, Last sequence update)
DE POSTERIOR SEX COMBS PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT "Drosophila genes Posterior Sex Combs and Suppressor two of zestle
RT encode proteins with homology to the murine bml-1 oncogene.";
RL Nature 353:351-353(1991).
CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC EXPRESSION PATTERNS OF THE HOMEOTIC SELECTOR GENES OF THE
CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X59275; CAA41965.1; -
DR PIR: S17983; S17983.
DR FLXBASE: FB9H005624; Psc.
DR INTERPRO: IPR001841; -
DR PFAM: PF00097; zf-C3HC4.1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4.1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 303 C3HC4-TYPE.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-GLY.
FT DOMAIN 1185 1189 POLY-PRO.
FT DOMAIN 1214 1217 POLY-PRO.
FT DOMAIN 1391 1396 POLY-PRO.
FT DOMAIN 1458 1461 POLY-GLY.
FT DOMAIN 1517 1520 POLY-ALA.
SQ SEQUENCE 1603 AA; 169999 MW; 77024F4097736473 CRC64;

Query Match 3.38; Score 128.5; DB 1; Length 1603;
Best Local Similarity 19.8%; Pred. No. 3.7;
Matches 164; Conservative 101; Mismatches 268; Indels 295; Gaps 39;

Oy 4 EVALALQAGSPSAANEAASQPADDEPLRRPRRDGGLGRSP-----GEPSAAVAPA 57
Db 130 ENALATTAALAAATTTTATPA-----LATSKAAKTIENGIRKKESTPPA 177

```

```

Oy 58 AAGCEAASAAAPALMRENAAGAAREAPATAVAGDNGSGLRREPADDEDDG 117
Db 178 VESVEASASSSSSSSSSSSSSSSSSSSSSTTRATSEDASSNGAAGA-----DEEK 224
Oy 118 EEEDEAAAAAAGIYRDNLITDGLTNGFHSGESDDORTSHASSSDWTTPRIGPY 177
Db 225 SEEDPTAAVAASS-----TATTSDLAT-----TSRRPYVL 256
Oy 178 TFVOQHLMIGDPTITLKDLPETIPPELDMITLQIVINITLSEPPKRRKKDINTIED 237
Db 257 TAVNPHITC-----HLQGYLTINATTIVECHLSCFCHSCILINL-----RKER----- 298
Oy 238 AVKLQEECKKIIVLTAGAVSVCIPDFRSRGIYARLAVDPDLPDQAMFIEYFRKD 297
Db 299 ---FCPCRMV-----NNAKPNIKSDTLTAIYKVLPGI-----YERLEMK- 339
Oy 298 PRPFKFAKE---IYFGQ-----FQP---SLCKFTAL---SDKEGKLLRNYT 336
Db 340 -FAFKDRPEEAAALATPEQRGDDTEHLIFSPDMSLSLEYALGELKLTDSPEL----- 393
Oy 337 QNIDTLEQVAGIQRILQCHGFATASCLIC-----KYKVDCAVAGDIFNOY----- 383
Db 394 -VDTLR-----PRYLQ-----PAMCRVSHLKKFVYDKFEIDAQFSDIMYKVKYTVL 441
Oy 384 -----VP-----RCPRCPADEPLAINKPEIVFGENLPQGFHRAKXY 420
Db 442 LDYITLMDIATYTWKRDAPMFRYRVESP--QPLVAPAPARV-----LPLAKLK-QER 493
Oy 421 DKDEVDLLIVIGSILKRVVAL-----IPSPHPEVP 452
Db 494 ENOEOLAVEVASS-KVEVVSIAEDOKAEASIKVEGESTREIVAEVIKDVATPTPTTL 552
Oy 453 QILINREPL-----PHLHFVYELGGDDVILNELC 482
Db 553 KLVINRNMIDKREKSHSPQLSSKSSKSSPCTPVSSPEPNKIKLIDISKONSVITIDMS 612
Oy 483 HRIAGEYAKLCNPKYLEITEKPPRQKELVHLSLPTPHISEDSSPPER----- 535
Db 613 DPERREIYK-----PLK-----PEKESRKK--DKDGSFKSSSSSSSGRRKKKSP 661
Oy 536 -TVP-----QDSVIATLVQATNNNVNDELVESE--SSCYEEKPQEVQTSRVNEN 582
Db 662 LVLPPLTIRTERIMSPSGVSTLSPVTSAGFSEDEPKSEIFALKPIKVKVESPERTLN 721
Oy 583 INVENDEKAVGSSSTADNERTVAETV-----RKWP-----NRLAKEOSKLEGNQ 631
Db 722 NEAITPSPSPVOQASAPSKSGNNLDDSLIMKPPSCMPKPSIASSKRKKEPV-KAVSRKQ 780
Oy 632 YLFPVNNYIFGAEVYSDSEDDVLSSSCGNSDSGTCQSPLEPL 679
Db 781 KLSPP-----LPTVDFKIRLPTVINGNS-SGTA-SRKIKPL 814

RESULT 14
HRX_HUMAN STANDARD: PRT: 3969 AA.
ID HRX_HUMAN 014845; Q16364; Q13743; Q13744; Q9DMA3;
AC 003164; 014845; Q16364; 013743; Q13744; Q9DMA3;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN HRX (ALL-1) (TRITHORAX-LIKE PROTEIN).
GN MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93046667; PubMed=1423624;
RA Tkachuk D.C., Kohler S., Cleary M.L.;
RT "Involvement of a homolog of Drosophila trithorax by 11q23
RT Chromosomal translocations in acute leukemias.";
Cell 71:691-700(1992).

```

[2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9629053; PubMed-8703835;
 RA Nilsson I., Loechner K., Siegler G., Grell J., Beck J.D., Fey G.H.,
 RT "Exon/Intron structure of the human ALL-1 (MLL) gene involved in
 translocations to chromosomal region 11q23 and acute leukaemias.";
 RL Br. J. Haematol. 93:966-972(1996).
 RN [3]
 RP SEQUENCE OF 1-1909 FROM N.A.
 RX MEDLINE-93390935; PubMed-8378076;
 RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
 RT Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
 "Two distinct portions of LTRG19/ENL at 19p13 are involved in t(11;19)
 leukemia.";
 RL Oncogene 8:2617-2625(1993).
 RN [4]
 RP SEQUENCE OF 1317-2328 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-93365134; PubMed-1303259;
 RA Djabbal M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
 RT "A trithorax-like gene is interrupted by chromosome 11q23
 translocations in acute leukaemias.";
 RL Nat. Genet. 2:113-118(1992).
 RN [5]
 RP SEQUENCE OF 1251-1538 FROM N.A.
 RX MEDLINE-94215165; PubMed-8162575;
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Cnaanl O.,
 RT Saito H., Croce C.M., Cnaanl E.;
 "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 involved in acute leukemia.";
 RL Cancer Res. 54:2326-2330(1994).
 RN [6]
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RX MEDLINE-9532025; PubMed-7598802;
 RA Mbangkolo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
 RT Rowley J.D., Diaz M.O.;
 "The human MLL gene: nucleotide sequence, homology to the Drosophila
 trx zinc-finger domain, and alternative splicing.";
 RL DNA Cell Biol. 14:475-485(1995).
 RN [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE-95315013; PubMed-794749;
 RA Marschalek R., Grell J., Loechner K., Nilsson I., Siegler G.,
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 transcripts in the acute lymphoblastic SEM cell line with chromosomal
 translocation t(4;11).";
 RL Br. J. Haematol. 90:308-320(1995).
 RN [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE-94020842; PubMed-841518;
 RA Forster A., Rabbitts T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 application to isolation of MLL fusion cDNAs from acute leukaemia
 translocations.";
 RL Oncogene 8:3157-3160(1993).
 RN [9]
 RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T LYMPHOCYTES.
 CC -1- DISEASE INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND ENL, T(4;11)(Q21;Q23)
 THAT INVOLVES MLL AND AFX4, T(9;11)(P22;Q23) THAT INVOLVES MLL AND
 AFX3, T(6;11)(Q27;Q23) THAT INVOLVES MLL AND AFX1, T(11;17)(Q23;Q21) THAT INVOLVES MLL
 AND AFX1, T(10;11)(P12;Q23) THAT INVOLVES MLL AND AFX10;
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND ENL AND T(11;11)(Q21;Q23)
 THAT INVOLVES MLL AND AFX10.
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
 FACTORS.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-FINGER DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L04284; AAAS6666.1; -;
 DR EMBL: 269744; CAA93625.1; -;
 DR EMBL: 269745; CAA93625.1; JOINED.
 DR EMBL: 269746; CAA93625.1; JOINED.
 DR EMBL: 269747; CAA93625.1; JOINED.
 DR EMBL: 269748; CAA93625.1; JOINED.
 DR EMBL: 269749; CAA93625.1; JOINED.
 DR EMBL: 269750; CAA93625.1; JOINED.
 DR EMBL: 269751; CAA93625.1; JOINED.
 DR EMBL: 269752; CAA93625.1; JOINED.
 DR EMBL: 269753; CAA93625.1; JOINED.
 DR EMBL: 269754; CAA93625.1; JOINED.
 DR EMBL: 269755; CAA93625.1; JOINED.
 DR EMBL: 269756; CAA93625.1; JOINED.
 DR EMBL: 269757; CAA93625.1; JOINED.
 DR EMBL: 269758; CAA93625.1; JOINED.
 DR EMBL: 269759; CAA93625.1; JOINED.
 DR EMBL: 269760; CAA93625.1; JOINED.
 DR EMBL: 269761; CAA93625.1; JOINED.
 DR EMBL: 269762; CAA93625.1; JOINED.
 DR EMBL: 269763; CAA93625.1; JOINED.
 DR EMBL: 269764; CAA93625.1; JOINED.
 DR EMBL: 269765; CAA93625.1; JOINED.
 DR EMBL: 269766; CAA93625.1; JOINED.
 DR EMBL: 269767; CAA93625.1; JOINED.
 DR EMBL: 269768; CAA93625.1; JOINED.
 DR EMBL: 269769; CAA93625.1; JOINED.
 DR EMBL: 269770; CAA93625.1; JOINED.
 DR EMBL: 269771; CAA93625.1; JOINED.
 DR EMBL: 269772; CAA93625.1; JOINED.
 DR EMBL: 269773; CAA93625.1; JOINED.
 DR EMBL: 269774; CAA93625.1; JOINED.
 DR EMBL: 269775; CAA93625.1; JOINED.
 DR EMBL: 269776; CAA93625.1; JOINED.
 DR EMBL: 269777; CAA93625.1; JOINED.
 DR EMBL: 269778; CAA93625.1; JOINED.
 DR EMBL: 269779; CAA93625.1; JOINED.
 DR EMBL: 269780; CAA93625.1; JOINED.
 DR EMBL: D14540; BAA03407.1; -;
 DR EMBL: L01986; AAA92511.1; -;
 DR EMBL: U04737; AAA18644.1; -;
 DR EMBL: S78570; AAB34770.1; -;
 DR EMBL: X83604; CAA58584.1; -;
 DR EMBL: S66432; AAB28545.1; -;
 DR TRANSFAC: T02337; -;
 DR MIM: 159555; -;
 DR INTERPRO: IPR001214; -;
 DR INTERPRO: IPR001487; -;
 DR INTERPRO: IPR001965; -;
 DR INTERPRO: IPR002857; -;
 DR PFAM: PF00628; PHD; 3.
 DR PFAM: PF00856; SET; 1.
 DR PFAM: PF02008; ZF-CXHC; 1.
 DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
 CC KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;
 CC Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 CC Alternative splicing.
 FT DOMAIN 17 102 ALA/GLY/SER-RICH.
 FT DNA_BIND 169 180 A.T HOOK (BY SIMILARITY).
 FT 217 227 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 301 309 A.T HOOK (BY SIMILARITY).
 FT DOMAIN 1431 1482 PHD 1.
 FT DOMAIN 1484 1533 PHD 2.
 FT DOMAIN 1566 1627 PHD 3.
 FT 1703 1748 BROMODOMAIN (DIVERGENT).
 CC -----

```

FT DOMAIN 3840 3969 SET DOMAIN.
FT DOMAIN 137 143 POLY-GLY.
FT DOMAIN 561 564 POLY-PRO.
FT DOMAIN 568 571 POLY-PRO.
FT VARSPLIC 1407 1444 MISSING (IN ISOFORM 14P-18B).
FT CONFLICT 144 144 E -> ELTTOIPCSKRTKRGHIDKTEFFLLANSMCLN
      (IN REF. 2).
FT CONFLICT 317 379 GLINSLEKPKQYKREKTEGTPPLTKEDKTVVROSPRIK
      (IN REF. 2).
FT CONFLICT 317 379 VRIIPSSKRTDATTAKOLLOLA -> VSSLILNPKSPKSG
      KTRKHEHLKRIKOLISDALKALEGSLQGLFLQNGQMPPL
      SNSYRCO (IN REF. 1).
FT CONFLICT 556 556 Q -> E (IN REF. 2).
FT CONFLICT 1487 1487 R -> G (IN REF. 5).
FT CONFLICT 1603 1603 S -> SCGE (IN REF. 2 AND 7).
FT CONFLICT 1616 1616 S -> C (IN REF. 6).
FT CONFLICT 1937 1937 O -> H (IN REF. 4).
FT CONFLICT 2181 2181 P -> S (IN REF. 4).
FT CONFLICT 3718 3718 R -> G (IN REF. 2).
FT CONFLICT 3759 3759 N -> D (IN REF. 2).
FT CONFLICT 3813 3813 D -> G (IN REF. 2).
FT CONFLICT 3901 3901 R -> A (IN REF. 2).
SO SEQUENCE 3969 AA: 431882 MW: C609CAF082E36CC9 CRC64;

```

Query Match 3.2%, Score 124.5; DB 1; Length 3969;
 Best Local Similarity 19.3%; Pred. No. 21;
 Matches 130; Conservative 92; Mismatches 234; Indels 217; Gaps 28;

```

QY 36 PRROPGIGRSGERSAVAAPAACE-----AASAAAPALMREAAGAAASEREP 88
DB 44 PPGVGGGGAPSPPAVAAAAAGSSGAGVPGAAASASSSSSSSSSSSSSSSSSASP 103
QY 89 ATAVAGDCDN-----GSGLRREPRADDFDEGEDEEAAMAAAAAIGYRDNL 139
DB 104 ALLRGPFGDALQYSAIGTLRFRFAVFGSGGGSGGSGDEQ----- 147
QY 140 LTDGLLNGFHCESDDDDT---SHAASSDWTFR-PRIGPYTVOQHLMTGDPRT-IL 194
DB 148 -----FLCFGSDDEVVRVSPRSPSVKTPRKRGRPRS-----GSDRNSAL 190
QY 195 KDLPETIPPELDMTLMQIVINILSEPRKKRKDINTIEDAVKLOECKIIVLGA 254
DB 191 SD--PSVSPPLKSE-----TKSGDIKIRKDSKSLT-----KKRGRPTFPGV 231
QY 255 GVSVSCGIPDFERSDGIYARLAVDPDLPDQAMFIDEFRKDPAPFFKFA---KEIYPG 311
DB 232 KIKITHG-----KDISELPKGKEDSLKIKRTPSATIQOAKIKKLAG 276
QY 312 QOPSLCHKFTALSDKEKLLRNTYQNTIDLEQVAGIQRILOCHGSFATASCLICKYVD 371
DB 277 KLSPLKSKF-----KTGKL-----QIG 293
QY 372 CEAVAGDIFNOVPRCPGCPADEPLAIKPELVFGENLPQGFHAMKYDK-----EV 425
DB 294 RKGV-----QIVRRRGRPESTE--RIKTPSGLINSELEK--OKVRKDKGTPTLKE 343
QY 426 DLLIIGSLKYRPAALIPSSIPHEVPOILINREPLPHLFDVLLGCDVITNELCHR- 484
DB 344 DKTIVROSPRIKPVRIIIPSSKRTDA-----TIKKLLQRA 379
QY 485 -----LGGEYAKLCNPKV-----LSEITEKPRPQKELVHLSLPPPL 524
DB 380 KKAQAKTEKEAALQGRKRVKQVKNIQFIWVYSAISSRIITKPRRIEDEDYD-PI 438
QY 525 HISESSSPER--TYPO--DSSVIAIIVQATNNNNNDLEVSSSCVEEKPOVQVSRNVE 581
DB 439 KIARLESIPNRSFASPCSSSEKSSAQSQSSSSPSVDYS--TDSQASSEIQ 497
QY 582 NINVENPDKAV-----GSSADKNERTVAE-----TVRKCMPRRLKEQIS 624
DB 498 VLPERSDTPREYHPLRPTISQSPENESNDRRSRKRYSVERSGSRITTKLSTIQSAPQOT 557
QY 625 KRLEGNOLYFVR 637

```

```

DB 558 SSSPPPLTPPP 570
RESULT 15
ICP4_HSVB STANDARD: PRT: 1487 AA.
ID ICP4_HSVB
AC P28925;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4 (155 KDA IMMEDIATE-EARLY
  PROTEIN).
DE IE OR 64.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;
  Alphaherpesvirinae; Varicellovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -I- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
  OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
  OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -I- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -I- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
  PHOSPHORYLATION.
CC -I- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86664; AAB02515.1; -
DR EMBL: M86664; AAB02499.1; -
DR PIR: A36802; EDBEF6.
DR HSSP: P04002; IWFA.
DR TRANSFAC: T00401; -
KW Early protein; Transcription regulation; Trans-acting factor;
  DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 181 213 SER-RICH.
FT DOMAIN 922 931 ARG/LYS-RICH (BASIC).
SO SEQUENCE 1487 AA: 154868 MW: 1D4397838D03680D CRC64;

```

```

Query Match 3.2%, Score 124; DB 1; Length 1487;
Best Local Similarity 25.7%; Pred. No. 6.3;
Matches 45; Conservative 19; Mismatches 71; Indels 40; Gaps 5;
QY 8 ALQAAGSPSAAAMEAQAQADEPLRKRRPRRDGPGLGR-----SPGEPSAAVAAPA 58
DB 87 AAGSGCPPTPAQAQPAS-PAPSP-----APGLAMLMKWVSHVAPNGRRATGSSS 137
QY 59 AGCEAASAAAPALMREAAGAAASERAPATAVAGDNGSLRREPRADDFDDGGE 118
DB 138 PG--GGDAADPYALDSDTETGPGSPQPEFPSSASGSGSPAPRVRSISISSSSSSSMD 195
QY 119 EEDDEAAAAAAGYRDNLLITDGLLNGFHCSDDDDDRTSHASSSDDMPRPR 173
DB 196 EDDQADGAGASS-----SSSSSSDDSDSDGCGEETPRPR 230

```

Search completed: February 16, 2001, 10:52:25
 Job time: 528 sec

?

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:41:38 ; Search time 51.38 Seconds
(without alignments)
257.578 Million cell updates/sec

Title: US-09-461-580A-1

Perfect score: 3854

Sequence: 1 MADEVALALQAAGSPSAAAA.....NEAIRFRELTVNVPDSKS 737

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6.COMB.pep.*
4: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep.*
5: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135	3.5	1805	1 US-07-853-913-2	Sequence 2, Appl1
2	122	3.2	738	3 US-08-664-038A-3	Sequence 3, Appl1
3	121	3.1	595	1 US-08-425-069-4	Sequence 4, Appl1
4	121	3.1	595	2 US-08-317-84B-4	Sequence 4, Appl1
5	120	3.1	873	3 US-08-990-140-2	Sequence 2, Appl1
6	119	3.1	160	5 5273901-7	Patent No. 5273901
7	119	3.1	180	5 5482709-6	Patent No. 5482709
8	114	3.0	866	3 US-08-651-999A-1	Sequence 1, Appl1
9	113	2.9	335	2 US-08-405-175A-6	Sequence 6, Appl1
10	112.5	2.9	332	2 US-08-405-175A-5	Sequence 5, Appl1
11	112	2.9	584	3 US-08-729-416C-17	Sequence 17, Appl1
12	111.5	2.9	832	1 US-08-209-747-2	Sequence 2, Appl1
13	111.5	2.9	832	1 US-08-458-298-2	Sequence 2, Appl1
14	108	2.8	98	2 US-07-814-220-2	Sequence 2, Appl1
15	108	2.8	98	2 US-07-812-421-2	Sequence 2, Appl1
16	108	2.8	783	5 5231168-2	Patent No. 5231168
17	105.5	2.7	584	3 US-08-729-416C-7	Sequence 7, Appl1
18	105	2.7	309	2 US-08-405-175A-7	Sequence 7, Appl1
19	104.5	2.7	833	1 US-07-977-434-4	Sequence 4, Appl1
20	104.5	2.7	893	4 US-08-458-819-4	Sequence 4, Appl1
21	104.5	2.7	893	4 PCT-US91-07035-4	Sequence 4, Appl1
22	104.5	2.7	1317	2 US-09-083-521-7	Sequence 7, Appl1
23	104	2.7	1435	2 US-08-568-459A-4	Sequence 4, Appl1
24	104	2.7	1435	2 US-08-487-828B-4	Sequence 4, Appl1
25	103	2.7	599	3 US-08-556-419-22	Sequence 22, Appl1
26	103	2.7	629	3 US-08-556-419-23	Sequence 23, Appl1
27	103	2.7	1503	3 US-08-976-255-14	Sequence 14, Appl1
28	103	2.7	2441	1 US-08-194-468-2	Sequence 2, Appl1

ALIGNMENTS

29	103	2.7	2441	3 US-08-961-739-2	Sequence 2, Appl1
30	102.5	2.7	309	2 US-08-405-175A-8	Sequence 8, Appl1
31	102.5	2.7	651	1 US-08-431-080-24	Sequence 24, Appl1
32	102.5	2.7	651	2 US-08-938-534-24	Sequence 24, Appl1
33	102.5	2.7	1043	3 US-08-928-361B-30	Sequence 30, Appl1
34	102.5	2.7	1721	3 US-08-700-651-5	Sequence 5, Appl1
35	102.5	2.7	1721	3 US-08-928-361B-6	Sequence 6, Appl1
36	102	2.6	968	3 US-08-651-999A-7	Sequence 7, Appl1
37	102	2.6	1298	2 US-08-690-473-2	Sequence 2, Appl1
38	102	2.6	1780	1 US-08-769-309A-5	Sequence 5, Appl1
39	102	2.6	1780	3 US-08-994-570-5	Sequence 5, Appl1
40	101.5	2.6	2860	2 US-08-826-267-2	Sequence 2, Appl1
41	101	2.6	281	2 US-08-405-175A-9	Sequence 9, Appl1
42	101	2.6	754	3 US-09-214-564A-2	Sequence 2, Appl1
43	100.5	2.6	731	2 US-08-911-364-1	Sequence 1, Appl1
44	100.5	2.6	792	2 US-08-678-039A-40	Sequence 40, Appl1
45	100	2.6	1514	2 US-08-853-310-4	Sequence 4, Appl1

RESULT 1
US-07-853-913-2
Sequence 2, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
TITLE OF INVENTION: Nestin Expression As An Indicator of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-913-2

Query Match 3.5%; Score 135; DB 1; Length 1805;
 Best Local Similarity 19.7%; Pred. No. 0.0063;
 Matches 155; Conservative 99; Mismatches 262; Indels 272; Gaps 36;

```

OY 41 PGLGRSPGSPSAVAAPAAAGCAASAAPALMREAAGAAASA---ERRAPA-----89
DB 392 PPISEAPCPNNAEVRQOEVLPLSLQQAPEPLMLKATVPSSAIIPELEPGKGQHPP 451
OY 90 -----TAVAGDDNGSGLRREPRADDPDDEGE-----EDEAAAAA 127
DB 452 DDLTSLATNLNPHHPTLEKDESS-----ESRVSSIFQEDGQIWELEKADIEVKE 506
OY 128 AAAATGYPNNLLTDLGLTNGFHCESDDDDRTSHASSSDMTPRPRIGYTFVOQHLM--185
DB 507 NSSAKTQESGIDTE-----ETQDSQ-----GP---LQKETLAA 537
OY 186 ICTDPTTIKDLLPETIPPELDMTLWQIVINILSEPPKR-----KRDII---NTI 235
DB 538 LGEELPLMSLKIQNYETAGKENCNSST--BGHLGTLEGPEKEKOIPLKSLKEKVESEKTL 595
OY 236 EDAAVLLQCKKTIYLTGAGVSCGIPDFRSRDGIYARLAVDPPLPQAMFIEYR 295
DB 596 ENGVPVLSL-----LKEEDTRTD-----OELMSPKGTL-----625
OY 296 KDPPEPFKAKEIYVQFQPSLCHKFIALSDREKGLLRNYTONIDTLEQVAGIQRILQCH 355
DB 626 ---KRRSLGKE-----SOEVVAPRSKEGN-LESMTAFKESQIPLG-----662
OY 356 GSFAPASCLICKYKDCAEAVRDIFNQVVRPCRPADPEPLAIKREIVFGENIPEQPH 415
DB 663 --FPAEDQMLERLVEKED-----QSEPRSPREEDQACRP-----LQKENQ 702
OY 416 RAMKXDKDE---VDLLIYVSSLKVRPALIIPSSIPHEVPLL--INREPLHLHVDVL 470
DB 703 EPLGVEAAGQILLERLIEKESQESLMS---PEEDQEGRSIQKENQPELGYEEAE---755
OY 471 LGDCVITINELCHRLGGEYAKLCNPFVLSLSEITEKPPRP--QKELVHLSLPT-----522
DB 756 ---DOMLERLLEKESQESLK---SPEENORIGKPLERENQSLKLEENQETFPVLESR 808
OY 523 -----PLH-----ISEDSSPERTVPODSSVIATLVDAQTN 554
DB 809 NORPLRSLVEEERQIVAPLEKVSQDSLSGLAEENVQPLRYLEDDCINKSLLEDKTKH 868
OY 555 NNNDLE-----VESSCIVEKPEQVOTSRNVENINENPDKAVSGSTADKNR 603
DB 869 SLGSLEDRNGDSIIIPQESETO-VSLRPEEDQRIVNHLEKESOE-----SR 916
OY 604 TSVAVETVRKCMFNRLAKEQISKR-LEGNQYLFVPPRYTFHGAEYVSDSEDDVLSSSCG 662
DB 917 SSEE-----EGYMERSLEGEN-----HSLSSVEKEDQVQSLK 953
OY 663 SUSDGTCOSPLEBLEPDESEIEEFYNGLEDTERPECCAGSGGAGADGDEVNEAIA 722
DB 954 ESQDSG-----KSLDES--QETFGPLEKENAE-----SLRSIAGQDDEQKLEGE 997
OY 723 TROELTDV 730
DB 998 TQOTLRAV 1005

```

RESULT 2

US-08-864-038A-3
 ; Sequence 3, Application US/08864038A
 ; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: KUNIO NAKASHIMA et al.
 ; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 ; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 ; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 ; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

```

; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEO ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinetada fucata.
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
; US-08-864-038A-3

```

Query Match 3.2%; Score 122; DB 3; Length 738;
 Best Local Similarity 32.3%; Pred. No. 0.022;
 Matches 61; Conservative 9; Mismatches 65; Indels 54; Gaps 8;

```

OY 5 VALALQAG-----SPSAAAMEAASQPADEPLRRPRRDGPGICR 45
DB 481 LAALAAAGAGGGGFGGLGGLGGSSAAAAAAGCGGRALRRALRROMRGGS 540
OY 46 SPGEPSAAVAPA-----AAGCEAASAAAPALMREAAGAA 81
DB 541 AAAAAAAGAGGGGCGGFGVGLGGFGGFGGSSAAAAAAGCGG 596
OY 82 SAERAPATAVAGDNGSGRLRREPRAADPDDEGEDEEAAAAAAGIYRNLILT 141
DB 597 GG-RRGGRGRGGDDP-GNGASAVAAAAAAGGSADVAIAAAAAAM-YGDS---A 650
OY 142 DGL-LTNGF 149
DB 651 DGPDFDNGF 659

```

RESULT 3

US-08-425-069-4
 ; Sequence 4, Application US/08425069
 ; Patent No. 5728810

; GENERAL INFORMATION:

; APPLICANT: Lewis, Randolph V.
 ; APPLICANT: Xu, Ming
 ; APPLICANT: Hinman, Michael B.
 ; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK

[illegible]

	Query Match	Similarity	Best Local Similarity	Matches	45;
	3.1%	Score 121; DB 2; Length 595;	34.9%; Pred. No. 0.019;	8; Mismatches 64; Indels 12; Gaps 4;	
QY	12 AGSPSAAMEAASG-----PADE-PLRKRRRDGPGLGRSGPSSAAVAPAAAGCEAA	64	: :		
Dd	308 AGSAAAAAAGPGGGGLGYGGQGPGGYGGQQGPG-GYGPGSASAAAAAAGPGQQQP	366	: :		
QY	65 SAAPALWREAAAGAASAEREPATVAAGCGDNGSGLRREPRAADDFPDDEGEEDENA	124	: ::: :		
Dd	367 GGYGGOOGPGSQPGSASA-----AAAAAAGPYGGYGPOOQGGPGGYAPGOOGPSGPSASNA	422	:		
QY	125 AAAAAAATG	133			
Dd	423 AAAAAAGPG	431			
RESULT 5					
US-08-990-140-2					
Sequence 2, Application US/08990140A					
Patent No. 6093795					
GENERAL INFORMATION:					
APPPLICANT: Olsen, Henrik S.					
APPPLICANT: Ruben, Steven M.					
APPPLICANT: Sonnenberg, Nahum					
APPPLICANT: Methot, Nathalie					
APPPLICANT: Rom, Efran					
TITLE OF INVENTION: Human PtiI-like Subunit Protein (hpriI) and Human					
TITLE OF INVENTION: eIF4GF-Ilike Protein (p97) Genes					
FILE REFERENCE: 1488..0700001					
CURRENT APPLICATION NUMBER: US/08/990.140A					
CURRENT FILING DATE: 1997-12-12					
EARLIER APPLICATION NUMBER: US'60/033,151					
EARLIER FILING DATE: 1996-12-13					
NUMBER OF SEQ ID NOS: 13					
SOFTWARE: PatentIn Ver. 2.1					
SEQ ID NO 2					
LENGTH: 873					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-08-990-140-2					
Query Match					
Best Local Similarity 3.1% Score 120; DB 3; Length 873;					
Matches 79; Conservative 41; Mismatches 113; Indels 126; Gaps 19;					
Y	2 ADEVALAQAAAGSPSA---- <td>57</td> <td></td> <td></td> <td></td>	57			

```
Db 52 SEEVGTA-EAGEPEVTEPAEAEASGSESP--SPPAAEELPQSHAPVPVPAAGEAEG 109
Oy 58 AAGCEAASAAALMREAGAAASAREAPATAVAGDGDNGSLRREPRAD--DEDD 115
Db 110 EGARDAGSRSRQAVSEDAAGNEGRAAEPRALENGDAD-----EFSFSDPEFVD 162
Oy 116 EGEER-----DEAAAAAAAIGYRDN-----LLTDLGLTNGRH 150
Db 163 VSEEEELGDLKDRPOEADGIDSVIYDNPQVGPDRLEKLNKHIKIPSKGKITNDREY 222
Oy 151 SCESEDDDDT-----SHA-----SSDWPTRPRRGPTFVOQH-----LMTGTP 190
Db 223 ---PEEDGKTGYIFLEVASPAHAYDANKAD-----GYKLDKQHTFRVNLFTDFDK 271
Oy 191 RTILKDL--LPETIPPELDMTLMQIYINILSEPPKRRKRDINTIEDAVKLTDECK-- 246
Db 272 YMTISDENDIPEKOPFKDGNLRVW-----LEEA-----ECRQ 305
Oy 247 -KIIVLTGAGVSVCGIPDRSRDGIYARLAVDFDLPDP---QAMFDIEYFRKDP 299
Db 306 YSVI-----FESGD---RTSIFWMDVKDPVSIERARWTEYVWRKSPK 345
```

RESULT 6

```
5273901-7
: Patent No. 5273901
: APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
: SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
: AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
: TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
: SPOZOITE 21.5 KB ANTIGEN, AC-6B
: NUMBER OF SEQUENCES: 11
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/581,693
: FILING DATE: 12-SEP-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 215,162
: FILING DATE: 05-JUL-1988
: APPLICATION NUMBER: 746,520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627,811
: FILING DATE: 05-JUL-1984
: SEQ ID NO: 7:
: LENGTH: 180
5273901-7
```

```
Query Match 3.1%; Score 119; DB 5; Length 180;
Best Local Similarity 33.6%; Pred. No. 0.004;
Matches 41; Conservative 8; Mismatches 41; Indels 32; Gaps 2;
```

```
Oy 10 QAAGSPSAAMEASQPADEPLKRRPRRDGPGIGRSPEPSAAVAPAAAGCEAASAAAP 69
Db 31 ERAAAPAAATAAPAAATA-----ATAATAATAATATPAAAAAP 69
Oy 70 AALMREAGAAASAREAPATAVAGDGDNGSLRREPRADDFDDDEGEDEEAAAAAA 129
Db 70 AAAAAAATAATGAAPAAATAAAGAAGAKAAAEART-----ETEGAGAGAA 118
Oy 130 AA 131
Db 119 EA 120
```

RESULT 7

```
5482709-6
: Patent No. 5482709
: APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
: SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
: AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
: TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
: ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
```

```
: NUMBER OF SEQUENCES: 10
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/148,432
: FILING DATE: 08-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 581,693
: FILING DATE: 12-SEP-1990
: APPLICATION NUMBER: 215,162
: FILING DATE: 05-JUL-1989
: APPLICATION NUMBER: 746,520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627,811
: FILING DATE: 05-JUL-1984
: SEQ ID NO: 6:
: LENGTH: 180
5482709-6
```

```
Query Match 3.1%; Score 119; DB 5; Length 180;
Best Local Similarity 33.6%; Pred. No. 0.004;
Matches 41; Conservative 8; Mismatches 41; Indels 32; Gaps 2;
```

```
Oy 10 QAAGSPSAAMEASQPADEPLKRRPRRDGPGIGRSPEPSAAVAPAAAGCEAASAAAP 69
Db 31 ERAAAPAAATAAPAAATA-----ATAATAATAATATPAAAAAP 69
Oy 70 AALMREAGAAASAREAPATAVAGDGDNGSLRREPRADDFDDDEGEDEEAAAAAA 129
Db 70 AAAAAAATAATGAAPAAATAAAGAAGAKAAAEART-----ETEGAGAGAA 118
Oy 130 AA 131
Db 119 EA 120
```

RESULT 8

```
US-08-651-999A-1
: Sequence 1, Application US/08651999A
: Patent No. 6031088
: GENERAL INFORMATION:
: APPLICANT: Stefan Somlo and Toshio Mochizuki
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
: STREET: 90 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10016
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
: MEDIUM TYPE: DISKETTE
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/651,999A
: FILING DATE: MAY 23, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: ELIZABETH A. BOGOSIAN
: REGISTRATION NUMBER: 39,911
: REFERENCE/DOCKET NUMBER: 96700/395
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-5995
: TELEFAX: (212) 286-0854 or 286-0082
: TELEX: TWX 710-581-4766
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 866 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
```

RESULT 9
 US-08-405-175A-6
 ; Sequence 6, Application US/08405175A
 ; Patent No. 5885772
 ; GENERAL INFORMATION:
 ; APPLICANT:ADEREM, ALAN A.
 ; APPLICANT:CHEN, JIAMIN
 ; APPLICANT:CHANG, SANDY
 ; TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KLAUBER & JACKSON
 ; STREET: 411 HACKENSACK AVENUE
 ; CITY: HACKENSACK
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/405,175A
 ; FILING DATE: 16-MAR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: JACKSON ESQ., DAVID A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-121A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201 487-5800
 ; TELEFAX: 201 343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:

RESULT 10
 US-08-405-175A-5
 Sequence 5, Application US/08405175A
 Patent No. 5885772
 GENERAL INFORMATION:
 APPLICANT: Aderem, Alan A.
 APPLICANT: Chen, Jiamin
 APPLICANT: Chang, Sandy
 TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/405,175A
 FILING DATE: 16-MAR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-121A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201 487-5800
 TELEFAX: 201 343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 332 amino acids
 TYPE: amino acid

Matches 42; Conservative 7; Mismatches 45; Indels 38; Gaps 2;

OY 2 ADEVALALQAGSPSAAMAEASQPADEPLKRRPRDGPGLGRSPGSESAVAAPAAAGC 61
Db 5 SDAAAAAATAATAAAAAATA-----TAAAAAATAA 38

OY 62 EAASAAAPALMREAAAGAAASAREAPATAVAGDNGSGLRREPRADDFDDDEGEED 121
Db 39 TAAAAAATAATAAAAAATAATAAAAAATAAKAALTAANAA----- 86

OY 122 EAAAAAAAIG 133
Db 87 AAAATAAAAARG 98

RESULT 15

US-07-812-421-2
; Sequence 2, Application US/07812421

; Patent No. 5932697

; GENERAL INFORMATION:

; APPLICANT: Caceci, Thomas

; APPLICANT: Szumanski, Maria B.W.

; TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESSES:

; ADDRESS: WHITHAM, CURTIS & WHITHAM

; STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,

; CITY: Reston

; STATE: VA

; COUNTRY: USA

; ZIP: 20191

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/812,421

; FILING DATE: 23-DEC-1991

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/588,437

; FILING DATE: 25-SEP-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitham, Michael E.

; REGISTRATION NUMBER: 32,635

; REFERENCE/DOCKET NUMBER: CIT.016

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-391-2510

; TELEFAX: 703-391-9035

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 98 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-812-421-2

Query Match 2.8%; Score 108; DB 2; Length 98;

Best Local Similarity 31.8%; Pred. No. 0.015; Mismatches 45; Indels 38; Gaps 2;

OY 2 ADEVALALQAGSPSAAMAEASQPADEPLKRRPRDGPGLGRSPGSESAVAAPAAAGC 61
Db 5 SDAAAAAATAATAAAAAATA-----TAAAAAATAA 38

OY 62 EAASAAAPALMREAAAGAAASAREAPATAVAGDNGSGLRREPRADDFDDDEGEED 121
Db 39 TAAAAAATAATAAAAAATAATAAAAAATAAKAALTAANAA----- 86

OY 122 EAAAAAAAIG 133
Db 87 AAAATAAAAARG 98

Search completed: February 16, 2001, 10:44:01
Job time: 143 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	385.4	100.0	737	11	Q9XG68	Q9xg68 mus musculus
2	331.3	86.0	747	4	Q9Y6F0	Q9y6f0 homo sapien
3	1169	30.3	823	5	Q9YK34	Q9yK34 drosophila
4	1157.5	30.0	823	5	Q96505	Q96505 drosophila
5	923.5	24.0	607	5	Q21921	Q21921 caenorhabdi
6	605.5	15.7	411	3	Q94640	Q94640 schizosacch
7	556.5	14.4	373	4	Q9Y6E9	Q9y6e9 homo sapien
8	555.5	14.4	352	4	Q9UNTO	Q9unto homo sapien
9	538	14.0	331	3	Q94066	Q94066 candida alb
10	513.5	13.3	332	3	Q9GUSN7	Q9gusn7 schizosacch
11	508	13.2	329	4	Q9Y6E8	Q9y6e8 homo sapien
12	493	12.8	254	4	Q95889	Q95889 homo sapien
13	406	10.5	351	5	Q96670	Q96670 trypanosoma
14	405.5	10.5	373	4	Q9P6C8	Q9p0c8 homo sapien
15	349	9.1	247	2	Q07595	Q07595 bacillus su
16	346	9.0	246	2	Q9WYW0	Q9wyw0 thermotoga
17	330.5	8.6	253	1	Q30124	Q30124 archaeoglob
18	295	7.7	245	1	Q28597	Q28597 archaeoglob
19	285.5	7.4	247	1	Q9YB13	Q9yb13 aeropyrum p

20	282.5	7.3	249	1	058669	058669	pyrococcus
21	267.5	6.9	250	1	090UE7	090ue7	pyrococcus
22	264	6.9	243	2	053700	053700	staphyloco
23	262	6.8	310	4	09y6E6	09y6e6	homo sapie
24	256	6.6	415	3	09UR39	09ur39	schizosacc
25	251.5	6.5	299	4	09NXA8	09nxa8	homo sapie
26	245	6.4	239	2	067919	067919	aquifex aq
27	237	6.1	314	4	09y6E7	09y6e7	homo sapie
28	233	6.0	326	4	043346	043346	homo sapie
29	230	6.0	237	2	09KPE5	09ke5	homo sapie
30	220	5.7	287	5	020480	020480	caenorhabd
31	216.5	5.6	299	2	09URJ35	09urj35	streptomy
32	216	5.6	377	4	09UOD1	09ud1	homo sapie
33	212	5.5	355	4	09NRC7	09nc7	homo sapie
34	211	5.5	237	2	006549	006549	mycobacter
35	210	5.4	259	2	09KRX4*	09krx4	vibrio crio
36	210	5.4	279	2	P75960	P75960	eschერი
37	199.5	5.2	208	2	09ZAB8	09zab8	actinobact
38	197.5	5.1	234	2	09ZJW8	09zjw8	heliobact
39	190	4.9	287	5	020481	020481	caenorhabd
40	187.5	4.9	233	2	09PNP1	09pnp1	campyloba
41	186.5	4.8	283	10	09XGT1	09xgt1	oryza sat
42	186	4.8	427	2	09JMN5	09jmn5	campyloba
43	186	4.8	3534	12	039266	039266	equine he
44	185.5	4.8	246	2	09RID4	09rid4	deinococcu
45	185	4.8	320	5	09UIA6	09uia6	leishmania

ALIGNMENTS

RESULT	1	
Q90XG8		
ID	Q90XG8	PRELIMINARY; PRT; 737 AA.
AC	Q90XG8;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE	SIR2ALPHA PROTEIN.	
GN	SIR2ALPHA.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SWISS WEBSPER/NIH;	
RA	Imai S., Armstrong C.M., Guarente L.;	
RT	"Silencing and ageing protein Sir2 is an NAD-dependent histone	
RT	deacetylase.";	
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF214646; AAF24983.1; -	
DR	HSSP; P04002; IATP.	
DR	INTERPRO; IPR003000; -	
DR	PRAM; PR02146; SIR2; 1.	
Q0	SEQUENCE 737 AA; 80371 MW; 7F15625E29433119 CRC64;	

Query Match	100.0%;	Score 3854;	DB 11;	Length 737;
Best Local Similarity	100.0%;	Pred. No. 4.1e-263;		
Matches 737; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	MADEVALLQAAGSFAAAAMEAASQPADEFLRKRRPRDDGGCLRSQGEPSAAVAAPAAAG	60
Db	1	MADEVALLQAAGSFAAAAMEAASQPADEFLRKRRPRDDGGCLRSQGEPSAAVAAPAAAG	60
Qy	61	CEAASAAAPALMREAAAGAAASAREAPATVAAGDGGGSLRREPRAADFDDDEGEEE	120
Db	61	CEAASAAAPALMREAAAGAAASAREAPATVAAGDGGGSLRREPRAADFDDDEGEEE	120
Qy	121	DEAAAAAAAAAIGYRDNLILLTGILTNFHSCEGDDDDRTSSAASSMTPPRPIGPTTFV	180
Db	121	DEAAAAAAAAAIGYRDNLILLTGILTNFHSCEGDDDDRTSSAASSMTPPRPIGPTTFV	180

```

OY 181 OOHLMIGDPTLTKDLPETIPPELDMTLMQIVINIILSEPPKRRKKKDIINTIEDAVK 240
DB 181 OOHLMIGDPTLTKDLPETIPPELDMTLMQIVINIILSEPPKRRKKKDIINTIEDAVK 240
OY 241 LLOECKKIIIVLGAGVSVSCGIPDRSRDGIYARLAVDFPDLDPQAMDIYFRKDRPP 300
DB 241 LLOECKKIIIVLGAGVSVSCGIPDRSRDGIYARLAVDFPDLDPQAMDIYFRKDRPP 300
OY 301 FPKFAKEIYPGOPSLCKHFTALSDKEGKLLRNTQNTDLEQVAGIORIIOCHGSFAT 360
DB 301 FPKFAKEIYPGOPSLCKHFTALSDKEGKLLRNTQNTDLEQVAGIORIIOCHGSFAT 360
OY 361 ASCLCKKRVDCAVRGDIFNOVPRPCPCPADEPLAIKPEIYFEGENIPEQFHRAMKY 420
DB 361 ASCLCKKRVDCAVRGDIFNOVPRPCPCPADEPLAIKPEIYFEGENIPEQFHRAMKY 420
OY 421 DKDEVDLIVTSSSLAKVRVALIPSSIPHEVQIILNREPLPHLHFDVELLDCDVIYNE 480
DB 421 DKDEVDLIVTSSSLAKVRVALIPSSIPHEVQIILNREPLPHLHFDVELLDCDVIYNE 480
OY 481 ICHRLGGEYAKLCNPKVLTSEITEKPPROKELVHLSLPTPLHISEDSSPERTVPD 540
DB 481 ICHRLGGEYAKLCNPKVLTSEITEKPPROKELVHLSLPTPLHISEDSSPERTVPD 540
OY 541 SSVIATLVDOATNNVNDLEVSSESSCVEEKPOEVQTSRNVENINVENPDKAVGSSTADK 600
DB 541 SSVIATLVDOATNNVNDLEVSSESSCVEEKPOEVQTSRNVENINVENPDKAVGSSTADK 600
OY 601 NERTVAEYVRKCMRNRAKEQISKRLEGNOYLFVPPNRYIFHGAEVYSDSEDDVLSSS 660
DB 601 NERTVAEYVRKCMRNRAKEQISKRLEGNOYLFVPPNRYIFHGAEVYSDSEDDVLSSS 660
OY 661 GGSNSDSGTCOSPSLEPDESEIEEFYNGLEDTEREPCAGSGSGFGADGDOEVYNA 720
DB 661 GGSNSDSGTCOSPSLEPDESEIEEFYNGLEDTEREPCAGSGSGFGADGDOEVYNA 720
OY 721 IATROELTDVNPSPDKS 737
DB 721 IATROELTDVNPSPDKS 737

```

RESULT 2

O9Y6FO PRELIMINARY; PRT; 747 AA.

AC O9Y6FO:

DT 01-NOV-1999 (Tremblrel. 12. Created)

DT 01-OCT-2000 (Tremblrel. 15. Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15. Last annotation update)

DE SIRTUIN TYPE 1.

GN SIRT1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCBI_TaxID=9606;

OX NCBI_TaxID=9606;

RA SEQUENCE FROM N.A.

RC TISSUE-TESTIS;

RX MEDLINE=99310604; PubMed=10381378;

RA FRYE R.A.;

RT "Characterization of Five Human cDNAs with Homology to the Yeast SIRT1 Gene; SIRT-Like Proteins (Sirtuins) Metabolize NAD and May Have Protein ADP-Ribosyltransferase Activity."

RL Blochem. Biophys. Res. Commun. 260:273-279(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-TESTIS;

RA FRYE R.A.;

DR Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF083106; AAD0849.2; -

DR INTERPRO: IPR003000; -

DR PFM; PFO2146; SIRT2; 1.

SO SEQUENCE 747 AA; 81680 MM; 203BEA6D73DA229F CRC64;

```

Query Match 86.0%; Score 3313; DB 4; Length 747;
Best Local Similarity 85.4%; Pred. No. 5e-225;
Matches 645; Conservative 35; Mismatches 45; Indels 26; Gaps 10;

```

OY 1 MADEVALALQAAAGSPSAAA-MEASQPADEPLRRKRRRDRGGLGRSPGPSAA-----V 54

DB 1 MADEVALALQPGSPSAAAGADREASSPAGEPLRRKRRRDRGGLGRSPGPSAAAREV 60

OY 55 ARAAGCEPAASAAAPALMR--EAAGAAASAREAPATVADGNGSLR--REPRRA 109

DB 61 PAARGCCPGAAA--AAIMREAEEAAAAGGQEQATAAEGGNGDPLQGPSREPLA 117

OY 110 DDF----DDDEGEEDDEAAAAAAGYRDMLLTGLDGLTNGFSCSDDDDRSHASS 165

DB 118 DNLVDEDDDEGEEDDE-----AAAAAIGRDMLLTGLDGLTNGFSCSDDEDRSHASS 173

OY 166 SDWTPRRPRIGPYTVOOHLMTGTPRTILKDLPETIPPELDMTLMQIVINIILSEPPK 225

DB 174 SDWTPRRPRIGPYTVOOHLMTGTPRTILKDLPETIPPELDMTLMQIVINIILSEPPK 233

OY 226 RKRKRDINTIEDAVKLLQCKKIIVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDLDP 285

DB 234 RKRKRDINTIEDAVKLLQCKKIIVLTGAGVSVSCGIPDRSRDGIYARLAVDFPDLDP 293

OY 286 QAMFIEYFRKDRPPFPKFAKEIYPGOPSLCKHFTALSDKEGKLLRNTQNTDLEQV 345

DB 294 QAMFIEYFRKDRPPFPKFAKEIYPGOPSLCKHFTALSDKEGKLLRNTQNTDLEQV 353

OY 346 AGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRPCPCPADEPLAIKPEIYF 405

DB 354 AGIORILQCHGSFATASCLICKYKVDCEAVRGDIFNOVPRPCPCPADEPLAIKPEIYF 413

OY 406 FGENPPEQFHRAMKDKDEVLDLIYGSLLKVRPALIPSSIPHEVQIILNREPLPHLH 465

DB 414 FGENPPEQFHRAMKDKDEVLDLIYGSLLKVRPALIPSSIPHEVQIILNREPLPHLH 473

OY 466 FVVELLGDGDVITINELCHRLGGEYAKLCNPKVLTSEITEKPPROKELVHLSLPTPLH 525

DB 474 FVVELLGDGDVITINELCHRLGGEYAKLCNPKVLTSEITEKPPROKELVHLSLPTPLH 533

OY 526 ISEDSSEPERTVPDSSVATLVDOATNNVNDLEVSSES-SCVEEKPOEVQTSRNVENI- 583

DB 534 VESDSSPERTSPDSSVATLVLDQAASN--DDLIVSESKGCMEEKPOEVQTSRNVESIA 592

OY 584 -NVENPDRKAVGSSSTADKNERTSVAEYRACKCPNRAKEQISKRLEGNOYLFVPPNRYIF 642

DB 593 EGMENPDLKNGSSYGEKERTSVAGYRACKCPNRAKEQISKRLEGNOYLFVPPNRYIF 652

OY 643 HGAEYYSDESDVLTSSSGSGNSDGTGOSPSLEPDESEIEEFYNGLEDTEREPCA 702

DB 653 HGAEYYSDESDVLTSSSGSGNSDGTGOSPSLEPDESEIEEFYNGLEDTEREPCA 712

OY 703 GGSFGADGDOEVYNEAATROELTDVNPSPDKS 737

DB 713 GGAGFGTGDGDDDEALNEAISVQEVTDNMVPSNKS 747

RESULT 3

O9YK34 PRELIMINARY; PRT; 823 AA.

AC O9YK34:

DT 01-MAY-2000 (Tremblrel. 13. Created)

DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15. Last annotation update)

DE SIRT2 PROTEIN.

GN SIRT2.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-BERKELEY.
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G., Morten J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Artl J.F., Agbayani A., An H.-U., Andrews-Plannkoch C., Baldwin D.,
 RA Bielew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Beres P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003639; AAF53248.1; -;
 DR FLYBASE: FBgn0024291; SIR2.
 DR INTERPRO: IPR003000; -;
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 91836 MW; 3DDIEBA975A263D3 CRC64;

Query Match 30.3%; Score 1169; DB 5; Length 823;
 Best Local Similarity 38.8%; Pred. No. 4.9e-74;
 Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;

QY 57 AAGCEAASAAAPALWREA-----GAASAREAPATAVAGDNGSGLRREPR----- 107
 Db 42 ASTSTEAEEAEATATTEPATSSELACNGEITKTTLAAEEOEIGANLEHRTKNPTKS 101
 QY 108 AADDFDDEGEDEEAASAAAAAAGYRDNLITDGLITNGFHSCESDDDRTSHASS- 166
 Db 102 MGEDEDEDEEEDDEE-----EDD-----EGGITGSNEDDESSNCSSSV 145
 QY 167 --DWTTPRPRIGYTFVVOOHLMTGTPRTILKDLPE--TIPPELDMMTLMQIVINILSE 222
 Db 146 EPDMKLR-----WLGREFYTGVRPVQVYASIMPHATGLAGTDSVLMYDLAHLNE 198
 QY 223 PPKRRKKRDINTIEDAVKLLOECKKIIVLTGAGSVSGCIPDFRSRDCIYARLAVDPDL 282
 Db 199 PKRKNKLASVTFEDVIVSLVKKSKIIIVLTGAGSVSGCITDFSTNCTIYARLADHPDL 258
 QY 283 PDPQAMPDIEYFRDPPRFKFAEIIYPGQFQPSLCHKFTALSDKEGKLRLNNTQNDITL 342
 Db 259 PDPQAMPDINFKRDPFPFYKFAREIYGFGEFQSPCHRFIKMLTKGLRLNNTQNDITL 318

QY 343 EQVAGIQRILQCHGSFATASCLICKYKVDCEAVBGDFINQVPRCPRP----- 391
 Db 319 ERVAGIGRVVICHGHSFSTASTCKCRFKCNADALADIFAGRIIPVCPQCPKRESDVASV 378
 QY 392 --ADEPL-----AIKPEIVFEGENLPEQFHRAMKYDKDEVLLIIVIGSSLKVRVALI 443
 Db 379 AVTEELRLQVLVENGIMKPDIVFEGGLPDEHYTWATKDVCDDLLIIVIGSSLKVRVAHI 438
 QY 444 PSSITPHEPQOLLINREPLPHIFVELLDGDDVIYNELCHLGE-----YALCCNPVKLS 500
 Db 439 PSSIPATVPQOLLINREQLHLKFPVELLDGSDVIYNQICHLSDNDCCWRGLCCDESULT 498
 QY 501 EITKPPAPKQELVHLSLPTPHIISEDSSPER-----TVPDSSVIATLVYQA 551
 Db 499 ESKELMP-PENSNHLLH---HLHRRCSSESERQSLDPTDQISNSADVIILSAG 554
 QY 552 TNNVNDLEVSSESC-----VE-----EKPO-EVOT 576
 Db 555 TCSD-SGFESSTFGCKRSTAAEAARIKITDILVELNETTALSCRLGLEGPOTYES 613
 QY 577 SRNV-----ENIVNPDPEKAVGSTADKN---ERTSAEIVRKCPWRKLAKEDISKRL 628
 Db 614 YRLHLSIDSSKSGIEQCDNEATPSYRPSNLVQETKTVAPSLTPIPOGRKROTAARLQ 673
 QY 629 -GNQYLFVPPRNYTFHGAEV-----YSPSEDDVLS--SCGSN---SDSGTC 670
 Db 674 PGTYSHTNNTSYVFPQAQVWMDYSDDDDEERSINRHSDLFGVGNHYKDDDEDAC 733
 QY 671 Q-----SPSL-----EPLDES 683
 Db 734 DLNAVPLSPPLPSLEAIVTDIVNGSNEPLPNS 768
 RESULT 4
 ID 096505 PRELIMINARY; PRT; 823 AA.
 AC 096505;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-Oct-2000 (TREMblrel. 15, Last annotation update)
 DE SIR2.
 GN SIR2.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Astron S.U., Rine J.;
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF068758; AAC79684.1; -;
 DR INTERPRO: IPR003000; -;
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A46235E7F CRC64;

Query Match 30.0%; Score 1157.5; DB 5; Length 823;
 Best Local Similarity 39.3%; Pred. No. 3.1e-73;
 Matches 289; Conservative 102; Mismatches 210; Indels 135; Gaps 25;

QY 52 AAVPAAAGCAASAAAPALWREAAGAAASAREAPATAVAGDNGSGLRREPR----- 107
 Db 42 ASTSTEAEEAEATATTEPAT--SELAC-KANGELTKTTLAAREOEIGANLEHRTKNPTK 98
 QY 108 -AADDFDDEGEDEEAASAAAAAAGYRDNLITDGLITNGFHSCESDDDRTSHASS 166
 Db 99 SMGEDEDEEEDDEE-----EDD-----EGGITGSNEDDESSNCSSS 142
 QY 167 --DWTTPRPRIGYTFVVOOHLMTGTPRTILKDLPE--TIPPELDMMTLMQIVINILS 221
 Db 143 VEPDMKLR-----WLGREFYTGVRPVQVYASIMPHATGLADDTDSVLMYDLAHLN 195

Oy	222	EPPKRRKKDINTEDAVKLLOEKKITIVLGAQVSVSCGIPDERSDGIYARAAVFPD	281
Db	196	EPRKNRKLASANTDEDVSLVKKSOKIIVLTGAQVSVSCGIPDFRSNGIYARAHPD	255
Oy	282	LPPDAWFDIEYFRKDRPPEFFKAKEIYPG--GFPSLCHEFTALSDKEKLLNNTONT	339
Db	256	LPDPDAWFDINVFYKRDRPPEFKFAREIYFGFQFOPSPCHRFIMLETCKLLRNNTONI	315
Oy	340	DTLEOVAGIOIKLIICHGSFAATSLCICKKYKDCAVAGDI.FNQVVPKPCRP-----	391
Db	316	DTLEKVAGIQVIECHGSFSTASTCKCFKCNMAALADPADIFPAQIRPVCPQCPPKREQSYD	375
Oy	392	-----ADEPL-----AIIKKEPIYFGEENLPEDCFHRAAKDKDEVDLLIYGSSIAKVRP	440
Db	376	ASVAVTBEELNAQIVENGIMRKDVIYFFEGBLPDEXHTYMAWDKDVCDLLIYGSSIAKVRP	435
Oy	441	ALLPSSIDPEVPOILLINREP.LPHLFVDELIGDCCVIINELCHRLOGE---YAKLCNPV	497
Db	436	AHIPSISPAIVPOLILINEQHHLHKPFVELLGDSVDIITINDICHLSPNDNCWRLOCDES	495
Oy	498	KLSSETTEKPRPKOKELVLSLPTPLPHISEDs-----SSEPRTVPQDSVATATLDQA	551
Db	496	VLTESKELMP-PEHSNHHLHHHLHNHRHCSESERQSLDTDTQSISKNSADIYLISAG	554
Oy	552	TNNNVNDLEVSESSC-----VE-----EKRO-EVOQT	576
Db	555	TCSD-SGESSTFCGCRKSTAAEAALIERIKTDILVELNETTALSCDRLGLEGPQTVES	613
Oy	577	SRNV-----EMINVENPDPKAVGSTADKN--ERTSVAETVKRCWNRNLAKEQISRLE	628
Db	614	YRHLSIDSSKSGSGLCOCDNEATPYRPSNLVQETKTVAPSLTPIPQORCKRGTAAERLQ	673
Oy	629	-GNQYLEVPNRYTFHGAEV-----YSDESDDYVLSSS-----SCGSN--SPSGTC	670
Db	674	PGRFSYTHNTNNSYVFPGAQVFWMDYDDDDDEEBERSHNHNSDLFGVGHNYKODDEDAC	733
Oy	671	Q-----SPSLPEPLE	680
Db	734	DLNAVPLSPILLPHSLE	749
<hr/>			
RESULT	5		
ID	021921	PRELIMINARY:	PRT: 607 AA.
AC	021921;		
DT	01-NOV-1996 (TREMBLrel_01, Created)		
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel_15, Last annotation update)		
DE	RIIAB.4 PROTEIN.		
GN	R1IAB.4.		
OS	Caeonhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;		
OC	Rhabdilitidae; Pelododerinae; Caeonhabditis.		
OX	NCBI_TaxID=6239;		
LN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Cummings P.;		
RL	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=94150718; Pubmed=7906398;		
RA	Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,		
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,		
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,		
RA	Shaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Weinstock L., Wilkinson-Sproat J.J., Wohlman P.,		
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
XT	elegans .;		

RL Nature 368:32-38(1994).
DR EMBL; Z70310; CA94364.1; -.
DR INTERPRO: IPR003000; -.
DR PFM; PF02146; STR2; 1.
SQ SEQUENCE 607 AA; 68765 MW; 71C720BA84DC8BD CRC64;

Query Match	24.0%;	Score 923.5;	DB 5;	Length 607;
Best Local Similarity	36.5%;	Pred. NO. 6.2e-57;		
Matches 217;	Conservative 106;	Mismatches 186;	Indels 85;	Gaps 15;

[illegible]

DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 KM Hypothetical protein: Nuclear protein.
 SQ SEQUENCE 471 AA; 53013 MW; 899AB1DA16E86B CRC64;

Query Match 15.7%; Score 605.5; DB 3; Length 471;
 Best Local Similarity 34.7%; Pred. No. 1,1e-34;
 Matches 170; Conservative 70; Mismatches 143; Indels 107; Gaps 19;

QY 50 PSAVAAPAAAGCEAASAAAPALREBAAGAAASAREAPATAVAGDGDNGSGLRPREPAA 109
 DB 11 PVASVSSSSG-----SSSGASLVDIMCGSKETEDEEV----- 44
 QY 110 DDEPDDEGEDEDAAAAAAIGYRDNLITDGLITFGHSCSDDDDRTSASSDWT 169
 DB 45 ----DSDEMDK-----VDSLETENT-----SDDER-----SEM 72
 QY 170 PRPRIGPYT-FVQOHLIGTDP-RTILKDL--LPETIPPELDMTLMQIVNLS-EP 223
 DB 73 RYLASGYAKFLKELYLBELPVRSILKKGINLPSAL--BEFEDIDLPLKREVKREV 130
 QY 224 PKRRKRDINTEDAVKLLQECKIIVLTGAGVSCGIPDFRSRDGIYARLAVDPDL 283
 DB 131 ARRIKLHFMTFEDVNLKKAKNVVVLVAGISTSLGIDFRSDNGCFARLASH--GLS 188
 QY 284 DPQMPDIEYFRKDRPFKRAKIYR--GQFQPSLCHKFTALSDKEKILRNTQNTDT 341
 DB 189 EPSMEFIHFFRENPDELYFFAROLDPEPTNHYSPS--HAFRLLEKKKKLSTLETQNDN 246
 QY 342 LEQYAGI--ORILQCHGFATASCLICKYKVCDAVRGIDINQVAPRCPRC----- 390
 DB 247 LEKKTGLSDNKKITIOCHSFARATCIKCKHKYDGSLEVEDINQVSYCNCEGKRPPLKLR 306
 QY 391 -----PADEPLA--IMKEIVFEGENLEPQ-FHRAKMYDKDEVLLI 429
 DB 307 VGQKKKKHYFSDSDSESDLLAQPIKMPDITFFFGALPDSPFNKVGSELEETDLII 366
 QY 430 VIGSSLVPRPALIPSSIPHEVPQILINREPLPHLHPDVELLGD-CVYINELCHRLG-- 486
 DB 367 CIGISLKVAPVSELISVIPPPTQIYISRTFVRHTQDVNLSPCDVMVIEICKRAGWL 426
 QY 487 GEYAKLCCNP 496
 DB 427 NELQALCDLP 436

RESULT 7
 Q9Y6E9 PRELIMINARY; PRT; 373 AA.
 ID Q9Y6E9;
 AC Q9Y6E9;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE SIRTUIN TYPE 2.
 GN SIRT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=99310604; PubMed=10381378;
 RA Fyfe R.A.;
 RT "Characterization of five human cDNAs with homology to the yeast SIR2
 RT gene: Sir2-like proteins (sirtuins) metabolize NAD and may have
 RT protein ADP-ribosyltransferase activity";
 RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
 DR EMBL: AF083107; AAD40850.1; -
 DR INTERPRO: IPR003000; -
 SQ SEQUENCE 373 AA; 41497 MW; C0985F3D81D5D815 CRC64;

Query Match 14.4%; Score 556.5; DB 4; Length 373;
 Best Local Similarity 35.6%; Pred. No. 2.2e-31;
 Matches 135; Conservative 69; Mismatches 106; Indels 69; Gaps 12;

QY 206 ELDDMTLMQIVINILSEPPKRRKRDINTIEDAVKLLQ--ECKRIIVLTGAGVSCGIP 263
 DB 19 EADMDLENLRSQTLISGSQKERLDELTLGVARYMQSERCRVICYLVAGISTSGACIP 78
 QY 264 DFRS-RDGIYARLAVDPDLDPQAMPDIEYFRKDRPFKRAKIYRPGQFQPSLCHKET 322
 DB 79 DFRSPSTGLYDNL--EKYHLYPEAIFETISYFKHPEFFALAKELVYGFQKPTICHYFM 136
 QY 323 ALSDKEKILRNTQNTDTLQVAGIQR--ILQCHGFATASCL--ICKYVDCVAVAGD 378
 DB 137 RLKMDKGLLCYTONIDTLEAGIODEDVEAHGTYSCHVASCRHETPLSMKEK 196
 QY 379 IFNOVPRCPRCPADDEPLAIKPEIVFEGENLPEQFHAKMYKRDDEVLLIVIGSSLKVR 438
 DB 197 IFSEVTPKCEDCQ-----SLVKPDIVFGESLPAFSCMSQDPLKVDLLVMGTSILOVQ 251
 QY 439 PVALIPSSIPHEVPQILINREPL--PHL-----HEVDLLGDCDVI 477
 DB 252 PFASLISKAPLSTFRLINKKAGQSDPLGIMIGLGGMDFDSKKAVRDVAMLGECDOG 311
 QY 478 INELCHRLG-----GEYAKLCC-----NPVLSLETKEPPRQKELVHLSLP 520
 DB 312 CLALAEELGWMKELEDLVRRHASIDAGAGVNP-STSASPKSPPPAR----- 361
 QY 521 PTLPHISEDSSSPERTVPO 539
 DB 362 -----DEARTTEREKQ 373

RESULT 8
 Q9UNT0 PRELIMINARY; PRT; 352 AA.
 ID Q9UNT0;
 AC Q9UNT0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE SILENCING INFORMATION REGULATOR 2-LIKE PROTEIN.
 GN SIRT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99321810; PubMed=10393250;
 RA Alshar G., Murnane J.P.;
 RT "Characterization of a human gene with sequence homology to
 RT Saccharomyces cerevisiae SIR2.";
 RL Gene 234:161-168(1999).
 DR EMBL: AF095714; AAD45971.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 352 AA; 39515 MW; FFED07DEF9E3416A CRC64;

Query Match 14.4%; Score 555.5; DB 4; Length 352;
 Best Local Similarity 36.2%; Pred. No. 2.4e-31;
 Matches 132; Conservative 68; Mismatches 96; Indels 69; Gaps 12;

QY 220 LSEPPKRRKRDINTIEDAVKLLQ--ECKRIIVLTGAGVSCGIPDRS-RDGIYARLA 276
 DB 12 LSLDSQKRLDELTLGVARYMQSERCRVICYLVAGISTSGACIPDRS-STGLYDNL- 70
 QY 277 VDPELDLPQAMPDIEYFRKDRPFKRAKIYRPGQFQPSLCHKFTALSDKEKILRNYT 336
 DB 71 -EKYHLYPEAIFETISYFKHPEFFALAKELVYGFQKPTICHYFMRLKDKGLLCRYT 129
 QY 337 QNIDTLEVAGIQR--ILQCHGFATASCL--ICKYKVCDAVRGIDINQVAPRCPRCPA 392

```

Db 130 QNIDTLERLAGLEQDVLVAHGTFTSHCVASACRHEYPPLSMKKERIFSEVTPKCEDCO- 188
OY 393 DEPLAIKMEPEIFEGENLPEQFHRAMKYKDEVDLLIVIGSSLKVRPALIPSSIPHEVP 452
Db 169 ----SLVAPDIYFFGSESLPAREFSCMQSDFLKVDLLVMGTSLQVOPFASLSKAPLSTP 244
OY 453 QILINREPL----PHL-----HFDVELLGDCDVIINELCHRLG----- 486
Db 245 RLLINKKAGQSDPFLGIMGLGGMDPFSKAYRDVAMLGSCDQGLALAEILGKKREL 304
OY 487 -----GEYAKLCC-----NPVKLSEITEKPPROKELVHLSELPPPLHISEDSSSPE 534
Db 305 EDLVRREHNASIDAQSGAGVNP-STASAPKSKSPPAK-----DEAKTTE 347
OY 535 RTVPQ 539
Db 348 REKPO 352

```

RESULT 9

```

O94066 PRELIMINARY; PRT; 331 AA.
AC 094066:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE TRANSCRIPTION REGULATORY PROTEIN.
OS CA99C4.08C.
OC Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC anamorphic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Murphy L., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=97435544; PubMed=9290243;
RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT "A Candida albicans genome project: cosmid contigs, physical mapping,
and gene isolation."
RL Fungal Genet. Biol. 21:308-314(1997).
DR EMBL: AL033503; CAA22018.1;
DR INTERPRO: IPR003000;
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 331 AA; 37509 MW; 8748DFB4ED79D10 CRC64;

```

Query Match 14.0%; Score 538; DB 3; Length 331;
 Best Local Similarity 34.8%; Pred. No. 3.7e-30;
 Matches 128; Conservative 69; Mismatches 117; Indels 54; Gaps 11;

```

OY 205 PELDMDTLMOQIYINILSEPPKRRKRDINTJEDAVKLLQECKRIIVLGAGVSGCIPD 264
Db 2 PSLDI-----LKPVAEAK-----NGKRYFFFNAGISTGAGIDP 37
OY 265 FRSRD-GIYARLAVDPDLDPQAMFDIEYFRKDRPFPEKKEIYPCQOPSLCHKFLA 323
Db 38 FRSPGTGLYANLAK--LNLPEAAVAFIDIDFEKEDKPPYTLAEELYPGNFAPTKPHHEIK 95
OY 324 LSDKSGKLLRNTQNTIDTLEQVAGIO--RILOCHGSFATASLLCKKYVDCAVAGDIEN 381
Db 96 LLODQGSILKRYATONIDTLERLAGVEDKRYIVAHGSFASNHCVDCHKEKTETTLTKYMD 155

```

```

OY 382 QVPRCPRCPADPLAIKMEIYFEGENLPEQFHRAMKYKDEVDLLIVIGSSLKVRPA 441
Db 156 KIPSCQHC-----GYVAPDIYFFGSESLPAREFSCMQSDFLKVDLLVMGTSLQVOPFASLSKAPLSTP 244
OY 442 LIPSSIPHEVPQILINREPLPHLH-----DVELLGDCDVIINELCHRLGGEYAKLCCNP 496
Db 211 SLPEVANKKCLRLVNLKKEVGFKEPRKSDIALHDCDIAVEAKCTLLGLD----- 262
OY 497 VKLSEITEKPPROKELVHLSELPPPLHISEDSSSPERTVPOSSVIATLVD-QATNNN 555
Db 263 DKLNEYEK-----EKIKSKAETKETKMHIEDKLKEAHLKEDKH--TTKYDNKEKOND 316
OY 556 VNDLEYSE 563
Db 317 ANDKELEQ 324

```

RESULT 10

```

O9USN7 PRELIMINARY; PRT; 332 AA.
AC 09USN7:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE SIR2-LIKE TRANSCRIPTIONAL REGULATORY PROTEIN.
GN SPC132.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Aert R., Volckaert G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121807; CAB58129.1;
DR INTERPRO: IPR003000;
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 332 AA; 37915 MW; 936A08D240C8BCF CRC64;

```

Query Match 13.3%; Score 513.5; DB 3; Length 332;
 Best Local Similarity 35.8%; Pred. No. 2e-28;
 Matches 126; Conservative 63; Mismatches 112; Indels 51; Gaps 10;

```

OY 225 IEDAVKLLQD--CKKIIVLTGAGVSGCIPDRSRD-GIYARLAVDPDLDPQAMFDI 291
Db 15 LEKVASLKEGKVKRICVAVGAGISTAGIDPDRSPETGIIYNNL--QRFNLPYAAVFDL 72
OY 292 EYFRDPPREFFAKEIYPCQOPSLCHKFLASKEGKILRNTQNTIDTLEQVAGI--Q 349
Db 73 SYFRKNPRFYLAEHLMEPEKRYTYHYFILLDLKRLQCYQNTIDTLERLAGVDPK 132
OY 350 RILQCHGSFATASCLCKKYKVDCEAVRGDIFNQVVRPCRPCADPLAIKMEIYFEGEN 409
Db 133 ALIEAHGSFOYSRCIECYEMAEYEVACIMQKVPKCNCK-----GLIKMIVFYGCG 187
OY 410 LPEQHRAMKYKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQILINREP---LPHLHF 466
Db 188 LPMREFHEHEKDTKCDMALVIGTSLVHPFADLEIYPNKQORVLLINREPAQDGERKK 247
OY 467 DVELLGDCDVIINELCHRLG--GEYAKLCCNPVKLSEITEKPPROKELVHLSELPPPL 524
Db 248 DIMITLGDCDSQVRLCKLKGWDELEKLIDTSVE--TLTE-----ISLL----- 290
OY 525 HISEDSSSPERTVPOSSVIATLVDQATNNNVDLEVSSESSVEEKKPOEVOT 576
Db 291 -----SVDSPIFEKNASQKKDQNSVNPFTKIEEKKKDEV 325

```

RESULT 11
 O9Y6E8 PRELIMINARY; PRT; 399 AA.

RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 [5]

RN SEQUENCE FROM N.A.

RP STRAIN-168;

RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y14082; CAA74510.1; -;

DR EMBL: Z99109; CAB12804.1; -;

DR INTERPRO: IPR003000; -;

DR PFAM: PF02146; SIR2; 1.

KW Hypothetical protein.

SO SEQUENCE 247 AA; 27417 MW; 8BAD6F797D3A99F7 CRC64;

Query Match 9.1%; Score 349; DB 2; Length 247;

Best Local Similarity 32.7%; Pred. No. 5e-17;

Matches 87; Conservative 54; Mismatches 91; Indels 34; Gaps 9;

```

QY 235 IEDAVKLLQCKKIIVLTGAGVSVCIPDFRSRDGIYARLAVDPDLPQAMFDIEYF 294
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1 MEYFKSILHEAQRIVLTGAGMSTESGIPDFRSAGIWE-----DASRMEM-SDYF 53

QY 295 RKDRPFPEKFAKEY----PGQFQPSLCHKFIALSDKKGKLLRNTQNTIDTLEQVAGIQR 350
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 54 LSYRLFWPKFKELFQMKMSGSEFEPNEGHLLELEKQKQKQVDTQNTIDGLHKRAGSRH 113

QY 351 ILQCHGSFATASCLICKYKVDCEAVRGDIFNQVPRCPRPADPEPL--AIKKPEIVFGE 408
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 114 VYELHGSIQRTACPCAGARYLP---HLERVEPECTAAGNNGDICGTIVLKTIVLFGD 169

QY 409 NLPEQFHRAMKYDK-DEVDLLIVIGSSILKVRPALIP--SSIPHEVQILINREPLPLH 465
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 170 AV--MHFDTLYEKLDQADLLVIGTSLVAPARFVPEDASLIPGMKKVIINLEPT----- 222

QY 466 FDVELLDGDDVIINELCHRLGGEYAK 491
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 223 -----YCDSLFDMVIRHOKTGEFAR 241

```

Search completed: February 16, 2001, 10:51:32

Job time: 592 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:51:32 ; Search time 137.23 Seconds
(without alignments)
209.254 Million cell updates/sec

Title: US-09-461-580A-4
Perfect score: 1288
Sequence: 1 VINILSEPPKRRKRDINTI.....ALIPSSIPHEVPQILINREP 245

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1288	100.0	737	11	09QXG8	09QXG8 mus musculu
2	1286	99.8	747	4	09YGE0	09YGE0 homo sapien
3	883.5	68.6	823	5	09VK34	09VK34 drosophila
4	873.5	68.0	823	5	096505	096505 drosophila
5	731.5	56.8	607	5	021921	021921 caenorhabdi
6	537	41.7	352	4	09UNTO	09UNTO homo sapien
7	537	41.7	373	4	09YGE9	09YGE9 homo sapien
8	515	40.0	471	3	094640	094640 schizosacch
9	497.5	38.6	399	4	09YGE8	09YGE8 homo sapien
10	492	38.2	254	4	09Y889	09Y889 homo sapien
11	466	36.2	332	3	09USN7	09USN7 schizosacch
12	465	36.1	331	3	094066	094066 candida alb
13	404.5	31.4	351	5	096670	096670 trypanosoma
14	387	30.0	373	4	09PGC8	09PGC8 homo sapien
15	346	26.9	246	2	09WYU0	09WYU0 chermotoga
16	331.5	25.7	247	2	007595	007595 bacillus su
17	318	24.7	253	1	030124	030124 archaeoglob
18	287.5	22.3	245	1	028597	028597 archaeoglob
19	282.5	21.9	249	1	058669	058669 pyrococcus

20	277	21.5	247	1	09YB13	09YB13 aeropyrum p
21	267.5	20.8	250	1	09UZE7	09UZE7 pyrococcus
22	257.5	20.0	243	2	053700	053700 staphylococ
23	246	19.1	415	3	09UR39	09UR39 schizosacch
24	245.5	19.1	299	4	09NXA8	09NXA8 homo sapien
25	245.5	19.1	310	4	09YGE6	09YGE6 homo sapien
26	237.5	18.4	239	2	067919	067919 aquifex aeo
27	237	18.4	314	4	09YGE7	09YGE7 homo sapien
28	233	18.1	326	4	043346	043346 homo sapien
29	226.5	17.6	237	2	09KEE5	09KEE5 bacillus ha
30	220	17.1	287	5	020480	020480 caenorhabdi
31	216.5	16.8	299	2	09RL35	09RL35 streptomyce
32	210	16.3	259	2	09KRX4	09KRX4 vibrio chol
33	206	16.0	279	2	P75960	P75960 escherichia
34	201.5	15.6	237	2	006549	006549 mycobacteri
35	198.5	15.4	208	2	09ZAB8	09ZAB8 actinobact
36	197.5	15.3	234	2	09ZJW8	09ZJW8 helicobacte
37	190	14.8	287	5	020481	020481 caenorhabdi
38	187	14.5	233	2	09PNP1	09PNP1 campylobact
39	186	14.4	227	2	09JN05	09JN05 campylobact
40	185.5	14.4	246	2	09RYD4	09RYD4 delinococcus
41	183	14.2	483	10	09XGT1	09XGT1 oryza sativ
42	180.5	14.0	204	5	09W4A1	09W4A1 drosophila
43	172.5	13.4	91	5	09NMB2	09NMB2 leishmania
44	171.5	13.3	320	5	09UJA6	09UJA6 leishmania
45	167	13.0	377	4	09UQD1	09UQD1 homo sapien

ALIGNMENTS

RESULT	ID	Q9QXG8	PRELIMINARY;	PRT;	737 AA.
Q9QXG8	Q9QXG8	Q9QXG8			
AC	Q9QXG8	Q9QXG8			
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)			
DE	SIR2ALPHA PROTEIN.				
GN	SIR2ALPHA.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxId=10090;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	SPRAIN-SWISS WEBSTER/NIH;				
RA	Imai S., Armstrong C.M., Guarente L.;				
RT	"Silencing and ageing protein Sir2 is an NAD-dependent histone				
RT	deacetylase."				
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF214646; AAF24983.1;				
DR	HSSP; P04002; IATF.				
DR	INTERPRO: IPR003000; -				
DR	PRAM: PR02146; SIR2; 1.				
SQ	SEQUENCE 737 AA: 80371 MW: 7F15625E29433119 CRC64:				
Query Match	100.0%; Score 1288; DB 11; Length 737;				
Best local Similarity	100.0%; Pred. No. 1.5e-109;				
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	VINILSEPPKRRKRDINTIEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 60			
DB	216	VINILSEPPKRRKRDINTIEDAVKLLQECKKIIVLTGAGVSVSCGIPDRSRDGIYARL 275			
QY	61	AVDPDPDPQAMDIYFKRDPFFKFAKEIYPGQFQPSLCKFKFALSDKEKLLRNY 120			
DB	276	AVDPDPDPQAMDIYFKRDPFFKFAKEIYPGQFQPSLCKFKFALSDKEKLLRNY 335			
QY	121	TQNTDTEQVAGIQRILQCGSFATASCLCKYKVDEAVAGDIFNOVVRCPADP 180			
DB	336	TQNTDTEQVAGIQRILQCGSFATASCLCKYKVDEAVAGDIFNOVVRCPADP 395			

```

QY 181 LAIMKPEIVFGENLPQGFHRAKMDKDEVDLLIYGSSLKVRPALIPSSIPHEVPQIL 240
DB 396 LAIMKPEIVFGENLPQGFHRAKMDKDEVDLLIYGSSLKVRPALIPSSIPHEVPQIL 455
QY 241 INREP 245
DB 456 INREP 460

RESULT 2
QY9Y6FO PRELIMINARY; PRT; 747 AA.
AC QY9Y6FO;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE SIRTUN TYPE 1.
GN SIRT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=99310604; PubMed=10381378;
RA Frye R.A.;
RT "Characterization of Five Human CDNs with Homology to the Yeast SIRT
RT Gene; Sirt-Like Proteins (Sirtuins) Metabolize NAD and May Have
RT Protein ADP-Ribosyltransferase Activity."
RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Frye R.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083106; AAD40849.2; -
DR INTERPRO; IPR003000; -
DR PFM; PF02146; SIRT2; 1.
SO SEQUENCE 747 AA; 81680 MW; 2D3BEA6D73DA229F CRC64;

Query Match 99.8%; Score 1286; DB 4; Length 747;
Best Local Similarity 99.6%; Pred. No. 2,3e-109;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VINITSEPPKRRKRDINTIEDAVKLLQDECKIIVLTGAGVSVCIGIPDRSRDITARL 60
DB 224 VINITSEPPKRRKRDINTIEDAVKLLQDECKIIVLTGAGVSVCIGIPDRSRDITARL 283
QY 61 AVDPDLDPQAMFDIEYFRKDRPFKFAKEIYPGQFQPSLCHKFTALSDEKGLLRNY 120
DB 284 AVDPDLDPQAMFDIEYFRKDRPFKFAKEIYPGQFQPSLCHKFTALSDEKGLLRNY 343
QY 121 TONIDTLEOVAGIORTIIOCHGSFATASCLICKYKVDCAVRGDIJFNQVPRPCRPADP 180
DB 344 TONIDTLEOVAGIORTIIOCHGSFATASCLICKYKVDCAVRGDIJFNQVPRPCRPADP 403
QY 181 LAIMKPEIVFGENLPQGFHRAKMDKDEVDLLIYGSSLKVRPALIPSSIPHEVPQIL 240
DB 404 LAIMKPEIVFGENLPQGFHRAKMDKDEVDLLIYGSSLKVRPALIPSSIPHEVPQIL 463
QY 241 INREP 245
DB 464 INREP 468

RESULT 3
QY9YK34 PRELIMINARY; PRT; 823 AA.
AC QY9YK34;
DT 01-MAY-2000 (Tremblrel. 13, Created)

```

```

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE SIRT2 PROTEIN.
GN SIRT2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V., Blazer V., Blazer V.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavel J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AEO03639; AAF53248.1; -
DR FLYBASE; FBgn0024291; SIRT2.
DR INTERPRO; IPR003000; -
DR PFM; PF02146; SIRT2; 1.
SO SEQUENCE 823 AA; 91836 MW; 3DD1EBA975A263D3 CRC64;

Query Match 68.6%; Score 883.5; DB 5; Length 823;
Best Local Similarity 63.6%; Pred. No. 1.8e-72;
Matches 166; Conservative 38; Mismatches 38; Indels 19; Gaps 2;

QY 3 NITSEPPKRRKRDINTIEDAVKLLQDECKIIVLTGAGVSVCIGIPDRSRDITARLAV 62
DB 194 HLNEPKRRKRLASVNTFDVILSKKSKITIVLTGAGVSVCIGIPDRSRDITARLAV 253
QY 63 DEPDLPQAMFDIEYFRKDRPFKFAKEIYPGQFQPSLCHKFTALSDEKGLLRNYTQ 122
DB 254 DEPDLPQAMFDIEYFRKDRPFKFAKEIYPGQFQPSLCHKFTALSDEKGLLRNYTQ 313
QY 123 NIDTLEOVAGIORTIIOCHGSFATASCLICKYKVDCAVRGDIJFNQVPRPCRP----- 176
DB 314 NIDTLEOVAGIORTIIOCHGSFATASCLICKYKVDCAVRGDIJFNQVPRPCRP----- 373

```


QY 177 -----ADEPL-----AIKPEIVFGENLPEOFHRAKXDKDEVLLIYIGSSLEKVR 223
 DB 374 VDSASVATEEELRQLVENGIMKPDIVFGEGLPBEYHVTATDKDVCDDLIVIGSSLEKVR 433
 QY 224 PVALIPSSIPHEVPQILINRE 244
 DB 434 PVAHIPSSIPATVPQILINRE 454

RESULT 4
 ID 096505 PRELIMINARY; PRT: 823 AA.
 AC 096505;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE SIR2.
 GN SIR2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Preygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Astrom S.U., Rine J.
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF068758; AAC79684.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A46235E7F CRC64;

Query Match 68.0%; Score 875.5; DB 5; Length 823;
 Best Local Similarity 63.5%; Pred. No. 9.5e-72;
 Matches 167; Conservative 37; Mismatches 38; Indels 21; Gaps 3;

QY 3 NITSEPPKRRKRKRDINTEDAVKILQECKRTIIVTGAGVSVCGLPDRSDGIYARLAV 62
 DB 192 HLLNEPRKRNKLASVNFDDVLSLVKRSQKILIVLTGAGVSVCGLPDRSTNGIYARLAV 251
 QY 63 DFDLPDPOAMFDIEYFRKDRPFEKFAKEIYPG--QOPSLCHKFTALSDEKGLLRNY 120
 DB 252 DFDLPDPOAMFDINFRKDRPFEKFAKEIYPG--QOPSLCHKFTALSDEKGLLRNY 311
 QY 121 TONIDLEQVAGIQRILQCHGSFATASCLICKYVDCENAVGDIYFNOVPRCP-- 176
 DB 312 TONIDTLERVAGIORVIECHGSFSTASTCKRCFKCNADALRADIFAQRIPVCPQCP-- 371
 QY 177 -----ADEPL-----AIKPEIVFGENLPEOFHRAKXDKDEVLLIYIGSSLEK 221
 DB 372 QSDASVATEEELRQLVENGIMKPDIVFGEGLPBEYHVTATDKDVCDDLIVIGSSLEK 431
 QY 222 VRYVALIPSSIPHEVPQILINRE 244
 DB 432 VRYVALIPSSIPATVPQILINRE 454

RESULT 5
 ID 021921 PRELIMINARY; PRT: 607 AA.
 AC 021921;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE R11A8.4 PROTEIN.
 GN R11A8.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pseudocercariae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Cummings P.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z70310; CAA94364.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 607 AA; 68765 MW; 71C720BAB41DC88D CRC64;

Query Match 56.8%; Score 731.5; DB 5; Length 607;
 Best Local Similarity 57.0%; Pred. No. 9.8e-59;
 Matches 138; Conservative 39; Mismatches 60; Indels 5; Gaps 1;

QY 3 NITSEPPKRRKRKRDINTEDAVKILQECKRTIIVTGAGVSVCGLPDRSDGIYARLAV 62
 DB 118 DLERAPVROKLTWYNLSADAVELFKTKKHILVLTGAGVSVCGLPDRSDGIYARLAV 177
 QY 63 DFDLPDPOAMFDIEYFRKDRPFEKFAKEIYPG--QOPSLCHKFTALSDEKGLLRNY 122
 DB 178 DFDLPDPOAMFDINFRKDRPFEKFAKEIYPG--QOPSLCHKFTALSDEKGLLRNY 237
 QY 123 NIDPLEQVAGIQRILQCHGSFATASCLICKYVDCENAVGDIYFNOVPRCP-- 182
 DB 238 NIDPLEQVAGIQRILQCHGSFATASCLICKYVDCENAVGDIYFNOVPRCP-- 292
 QY 183 IMKPEIVFGENLPEOFHRAKXDKDEVLLIYIGSSLEKVRVALIPSSIPHEVPQILIN 242
 DB 293 VIKNIYVFEGLDIREFHQHTEDKHVYDLIVIGSSLEKVRVALIPCHVKNVQILIN 352
 QY 243 RE 244
 DB 353 RE 354

RESULT 6
 ID 09UNTO PRELIMINARY; PRT: 352 AA.
 AC 09UNTO;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE SILENCING INFORMATION REGULATOR 2-LIKE PROTEIN.
 GN SIR2L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99321810; PubMed=10393250;
 RA Afshar G., Murmane J.P.;
 RT "Characterization of a human gene with sequence homology to
 RT Saccharomyces cerevisiae SIR2.";
 RL Gene 234:161-168(1999).
 DR EMBL: AF095714; AAD45971.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.

SO SEQUENCE 352 AA; 39515 MW; FRED07DEF9E3416A CRC64;

Query Match 41.7%; Score 537; DB 4; Length 352;
Best Local Similarity 43.7%; Pred. No. 3.1e-41;
Matches 108; Conservative 59; Mismatches 66; Indels 14; Gaps 6;

OY 5 LSEPPKRRKNDINTEDAVKLLQ--ECKKIIVLTGAGVSCGIPDRS-RDGIYARLA 61
DB 12 LSLGSKERLDELTELEGVARYMOSERCRIVICLVAGISTAGIPDRSPSTGLYDNL- 70
OY 62 VDFPLPQAMFIDIEFRKDPREFKFAKEIYPOGQPSLCHKRIALSDKGLLRNT 121
DB 71 EKYLPPPEALFEISYFKRHPPEFALAKELYPGQFKPTICHYFMRLKDKGLLRCT 129
OY 122 QNIDPLEQVAGIQR--IIQCHGSFATASCL--ICKYKVDCEAVRGDIFNOVYPRCPRA 177
DB 130 QNIDPLEKRIAGLEQEDVLEAHGTFTSHCVASCHNEYPLSMKKERISFVTPKCEDCO- 188
OY 178 DEPLAIKPEIYFEGENLPEQFHRAMKYDKDEVDLLIYSSLSKYRVPVALLPSSIPHEVP 237
DB 189 ---SLVPRDIYFEGESLPARFFSCMSQDFLKVLLVGTSLQVQPFASLSKAPLSTP 244
OY 238 QILINRE 244
DB 245 RLINKE 251

RESULT 7

OY96E9 PRELIMINARY; PRT; 373 AA.
AC OY96E9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE SIRTUIN TYPE 2.
CN SIRT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=99310604; PubMed=10381378;
RA Frye R.A.;
RT "Characterization of five human cDNAs with homology to the yeast SIRT2 gene: SIRT2-like proteins (sirtuins) metabolize NAD and may have
RT protein ADP-ribosyltransferase activity.";
RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
DR EMBL; AF083107; AAD40850.1; -.
DR INTERPRO; IPR003000; -.
SO SEQUENCE 373 AA; 41497 MW; C0985F3D81D5D815 CRC64;

Query Match 41.7%; Score 537; DB 4; Length 373;
Best Local Similarity 43.7%; Pred. No. 3.3e-41;
Matches 108; Conservative 59; Mismatches 66; Indels 14; Gaps 6;

OY 5 LSEPPKRRKNDINTEDAVKLLQ--ECKKIIVLTGAGVSCGIPDRS-RDGIYARLA 61
DB 33 LSLGSKERLDELTELEGVARYMOSERCRIVICLVAGISTAGIPDRSPSTGLYDNL- 91
OY 62 VDFPLPQAMFIDIEFRKDPREFKFAKEIYPOGQPSLCHKRIALSDKGLLRNT 121
DB 92 EKYLPPPEALFEISYFKRHPPEFALAKELYPGQFKPTICHYFMRLKDKGLLRCT 150
OY 122 QNIDPLEQVAGIQR--IIQCHGSFATASCL--ICKYKVDCEAVRGDIFNOVYPRCPRA 177
DB 151 QNIDPLEKRIAGLEQEDVLEAHGTFTSHCVASCHNEYPLSMKKERISFVTPKCEDCO- 209
OY 178 DEPLAIKPEIYFEGENLPEQFHRAMKYDKDEVDLLIYSSLSKYRVPVALLPSSIPHEVP 237

DB 210 ---SLVPRDIYFEGESLPARFFSCMSQDFLKVLLVGTSLQVQPFASLSKAPLSTP 265
OY 238 QILINRE 244
DB 266 RLINKE 272

RESULT 8

OY9640 PRELIMINARY; PRT; 471 AA.
AC OY9640;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE REGULATORY PROTEIN C16D10.07C.
CN SPC16D10.07C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Brown D., Churcher C.M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO THE SIRT2 FAMILY.
DR EMBL; AL0355637; CAB85511.2; -.
DR INTERPRO; IPR003000; -.
DR PFAM; PF02146; SIRT2; 1.
KW Hypothetical protein; Nuclear protein.
SO SEQUENCE 471 AA; 53013 MW; 899AB11DA16E868B CRC64;

Query Match 40.0%; Score 515; DB 3; Length 471;
Best Local Similarity 43.6%; Pred. No. 4.6e-39;
Matches 119; Conservative 39; Mismatches 77; Indels 38; Gaps 7;

OY 7 EPPKRRKNDINTEDAVKLLQ--ECKKIIVLTGAGVSCGIPDRS-RDGIYARLA 66
DB 129 EVARRIKLPHEFTEDVYVNLKAKANNVYLVAGAGISTGLIDFSDNGFTARLARH--G 186
OY 67 LPDQAMFIDIEFRKDPREFKFAKEIY--GOQPSLCHKRIALSDKGLLRNTONI 124
DB 187 LSEPEMEDIHFTFENPFEIYTFARDLLPENHNSPS--HAFIRLEKKNNLSTLFTONI 244
OY 125 DTELEVAGI--QRIQCHGSFATASCLICKYKVDCEAVRGDIFNOVYPRCPRC----- 175
DB 245 DNLEKKTGLSDNKIIQCHGSFATATCICKYKVDSELYEDIRNORVSYCNECGKPKL 304
OY 176 -----PADEPLA---IKKPEIYFEGENLPEQ-FHRAMKYDKDEVDL 212
DB 305 RRVGONKKEKHYSFGDSESESDLAOPGIMKPDITTFGEALPDSFFKVGSGELEFIDL 364
OY 213 LIVIGSLKVRVPVALLPSSIPHEVPQILINRE 245
DB 365 LICICTSLKAVPVSLLISVIPPTPQIYISRP 397
RESULT 9
OY96E8 PRELIMINARY; PRT; 399 AA.
AC OY96E8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE SIRTUIN TYPE 3.
CN SIRT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

OS Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC anamorphic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;

RP SEQUENCE FROM N.A.
 RC STRAIN=161;
 RA Murphy L., Harris D.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=161;
 RA Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=161;
 RA MEDLINE=97435544; PubMed=9290243;
 RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
 RT "A Candida albicans genome project: cosmid contigs, physical mapping,
 and gene isolation";
 RL Fungal Genet. Biol. 21:308-314(1997).
 DR EMBL, AL033503, CAA22018.1,--.
 DR INTERPRO, IPR003000; -.
 DR PFAM, PF02146; SIR2, 1.
 SO SEQUENCE 331 AA; 37509 MW; 8748DFB84ED79D10.CRC64;

Query Match 36.1%; Score 465; DB 3; Length 331;
 Best Local Similarity 40.8%; Pred. No. 1.1e-34;
 Matches 95; Conservative 53; Mismatches 71; Indels 14; Gaps 5;

QY 19 TIEDAVKLLQEC---KTIIVLTGAGVSGIPDFRSRD-GIVARLAVDPDLPDPOAM 73
 DB 3 SLDDLKVALAEVKKKGVTFEFGNGISTGACIPDRSPDGLVNLAK--LNLFAEAV 60
 QY 74 FDIEYERKDPPEFEKFAEIVPGQFQPSLCKHFKTALSDKEKLLNNYQNTDLEQVAGI 133
 DB 61 FDIDFEKEDPPEFYTLAEELYPGNFAPTKFHFKLQDQSLKRVYQNTDLERLAGV 120
 QY 134 Q-RILQCHGSFATASCLCKKYKDCVAVRGDIENOVVPRPCPADPLAIKREIYFF 191
 DB 121 EDKXIVEHGSGFASNHCVDCHEKEMTEFLKTYMKDKKIPSOCHCE---GVYKPDIVFF 175
 QY 192 GENLPEQFHRAMKYDKDEVLLIVIGSSLKRPVALIPSSIPHEVQOLINRE 244
 DB 176 GEGLPVKFEDMEDDCEDVEVAIVAGISLIVPPFASLGEVKKKCLRVLVNKE 228

RESULT 13
 096670 PRELIMINARY; PRT: 351 AA.
 ID 096670;
 AC 096670;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE SIR2 HOMOLOG.
 OS Trypanosoma brucei.
 CC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=427;
 RA Hoek M., Wirtz E., Cross G.A.M.;
 RT "Cloning of a Trypanosoma brucei Sir2 homolog";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF102869; AAC73004.1; -.
 DR INTERPRO, IPR003000; -.
 DR PFAM, PF02146; SIR2, 1.
 SO SEQUENCE 351 AA; 38529 MW; 9EA0BD5106670575.CRC64;

Query Match 31.4%; Score 404.5; DB 5; Length 351;
 Best Local Similarity 36.9%; Pred. No. 4.1e-29;
 Matches 92; Conservative 49; Mismatches 83; Indels 25; Gaps 8;

QY 3 NILSEPPKRRKKRDINTIEDAVKLIQ--ECKIIVLTGAGVSGIPDFRS-RDGIYAR 59
 DB 10 HVGEP-----TFEGLARFIERNNITKIFVMVAGISVAGIDPDFRSPHTGLYAK 59
 QY 60 LAVDFPDLDPQAMFDIEYERKDPPEFKFA--KEIYGGQPSLCKHFKTALSDKEKLL 117
 DB 60 LS--RYNLNSPEDASLPLRQPSVFYNIIMDDMLWPKTCPTVHHFISLAKGMLL 117
 QY 118 RNYQNTDLEQVAGIQR--ILOCHGSFATASCLCKKYKDCVAVRGDIFNOVPRPCRC 175
 DB 118 CCGTQNTDGLERAGCIPESLLEAHGSGFASNVDCVCHAKYQNTINARATFRAGKYPHCNQ 177
 QY 176 PADEPLAIKPEIYFEGNLPQEFHRAMKYDKDEVLLIVIGSSLKRPVALIPSSIPHE 235
 DB 178 G----GIVKPDVVFEGNLPQEFHRAMKYDKDEVLLIVIGSSLKRPVALIPSSIPHE 235
 QY 236 VPOLINRE 244
 DB 232 VPRVLEMLE 240

RESULT 14
 09P0G8 PRELIMINARY; PRT: 373 AA.
 AC 09P0G8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE SILENT INFORMATION REGULATOR 2 HOMOLOG.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADRENAL GLAND;
 RA Ren S., Fu S., Gu J., Jin W., Gu Y., Huang Q., Dong H., Yu Y., Fu G.,
 RA Wang Y., Chen Z., Han Z.;
 RT "A novel gene expressed in human adrenal gland";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADRENAL GLAND;
 RA Peng Y., Gu Y., Li Y., Fu S., Gu J., Zhang L., Jiang C., Yu Y.,
 RA Han Z., Wang Y., Chen Z., Fu G.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF160214; AAF67015.1; -.
 SO SEQUENCE 373 AA; 40737 MW; 81EAA55453F3638F.CRC64;

Query Match 30.0%; Score 387; DB 4; Length 373;
 Best Local Similarity 35.4%; Pred. No. 1.8e-27;
 Matches 93; Conservative 52; Mismatches 72; Indels 46; Gaps 9;

QY 5 LSEPPKRRKKRDINTIEDAVKLIQ--ECKIIVLTGAGVSGIPDFRS-RDGIYARLA 61
 DB 33 LLSGQKRELRDELDELLEAVARYMQSERCRVILCVAGISFASGIPDRSPSTGLYDLN- 91
 QY 62 VDFPDLDPQAMFDIEYERKDPPEFKFAKITYGGQF-----PSLC- 103
 DB 92 EKXHLPRPEALFEISYFKKHPPEFFALAKELYGQFQANHLSELHAPAEQOGATPALH 150
 QY 104 --HKFIALSDKEKLLRNYQNTDLEQVAGIQRILQCHGSFATASCLCKKYKDCVAVR 161
 DB 151 AEHRYPGANSRAG-----TGLGGGARHLL--HTLRAN---CHHEIFLSMK 194
 QY 162 GDIFNOVPRPCPADPLAIKPEIYFEGNLPQEFHRAMKYDKDEVLLIVIGSSLK 221
 DB 195 EKIRSEVTPKCEDCQ-----SLVKPDIVFEGESLPARFSCMQSDPLKVDLLVMGSLQ 249

RESULT	15	
Q9MYW0		
Q9MYW0	PRELIMINARY:	PRT: 246 AA.
AC Q9MYW0		
DT 01-NOV-1999	(TIREMBLrel. 12, Created)	
DT 01-NOV-1999	(TIREMBLrel. 12, Last sequence update)	
DT 01-JUN-2000	(TIREMBLrel. 14, Last annotation update)	
DE REGULATORY PROTEIN, SIR2 FAMILY.		
GN TM0490		

Query Match	25.9%	Score 346	DB 2	Length 246
Best Local Similarity	37.3%	Pred. No. 6e-24		
Matches 79; Conservative	41;	Mismatches	76;	Indels 16; Gaps 4;

Search completed: February 16, 2001, 10:51:34
Job time: 594 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:51:34 ; Search time 137.23 Seconds
(without alignments)
109.325 Million cell updates/sec

Title: US-09-461-580A-9
Perfect score: 682
Sequence: 1 IVLTGAGVSVSCGIPDRSR.....SFATASCLICKYKDCFAVR 128

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rentent:*
13: sp.virus:*
14: sp.vertebrate:*
15: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	100.0	737	11	O9GXG8
2	680	99.7	747	4	O9Y6F0
3	550	80.6	823	5	O9YK34
4	542	79.5	823	5	O96505
5	465	68.2	607	5	O21921
6	327.5	48.0	254	4	O95889
7	327.5	48.0	352	4	O9UNTO
8	327.5	48.0	373	4	O9Y6E9
9	315.5	46.3	331	3	O94066
10	311.5	45.7	399	4	O9Y6E8
11	309	45.3	471	3	O94640
12	281.5	41.3	332	3	O9USN7
13	242.5	35.6	351	5	O96670
14	230.5	33.8	246	2	O9YRW0
15	218.5	32.0	247	2	O9Y759
16	207.5	30.4	253	4	O30124
17	193	28.3	373	4	O9P0G8
18	185	27.1	247	1	O9YB13
19	181	26.5	245	1	O28597

20	175.5	25.7	415	3	O9UR39	O9UR39
21	174	25.5	243	2	O53700	O53700
22	166.5	24.4	204	5	O9W4A1	O9W4A1
23	160	23.5	249	1	O58669	O58669
24	159	23.3	239	2	O67919	O67919
25	153	22.4	299	4	O9NXX8	O9NXX8
26	153	22.4	310	4	O9Y6E6	O9Y6E6
27	152	22.3	250	1	O9UZE7	O9UZE7
28	152	22.3	299	1	O9URJ5	O9URJ5
29	151.5	22.2	287	5	O20480	O20480
30	151.5	22.2	314	4	O9Y6E7	O9Y6E7
31	149.5	21.9	259	2	O9KRX4	O9KRX4
32	147.5	21.6	326	4	O43346	O43346
33	144	21.1	91	5	O9NMB2	O9NMB2
34	143.5	21.0	233	3	O9PVP1	O9PVP1
35	143.5	21.0	279	2	P75960	P75960
36	142	20.8	237	2	O9KEE5	O9KEE5
37	141.5	20.7	287	5	O20481	O20481
38	133	19.5	208	2	O9ZAB8	O9ZAB8
39	131.5	19.3	227	2	O9JN05	O9JN05
40	126.5	18.5	234	2	O9ZCW8	O9ZCW8
41	122	17.9	246	2	O9RYD4	O9RYD4
42	119	17.4	483	10	O9XGT1	O9XGT1
43	115	16.9	1095	5	O01478	O01478
44	113.5	16.6	317	5	O9VH08	O9VH08
45	111	16.3	237	2	O06549	O06549

ALIGNMENTS

RESULT 1
ID O9GXG8 PRELIMINARY; PRT; 737 AA.
AC O9GXG8;
DT 01-MAY-2000 (TREMUREL. 13, Created)
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
DT 01-OCT-2000 (TREMUREL. 15, Last annotation update)
DE SIR2ALPHA PROTEIN.
GN SIR2ALPHA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIH;
RA Imai S., Armstrong C.M., Guarente L.,
RT "Silencing and ageing protein Sir2 is an NAD-dependent histone
RT deacetylase.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF214646; AAF24983.1; -
DR HSSP: P04002; IATF.
DR INTERPRO: IPR003000; -
DR PRAM: PR02146; SIR2; 1.
SQ SEQUENCE 737 AA; 80371 MW; 7E15625E29433119 CRC64;

Query Match 100.0%; Score 682; DB 11; Length 737;
Best local Similarity 100.0%; Pred No. 6.1e-65;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	IVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFIDYFRKDRPFKFAKEI 60
DB	249	IVLTGAGVSVSCGIPDRSRDGIYARLAVDPDLPDQAMFIDYFRKDRPFKFAKEI 308
OY	61	YPOQFQPSLCHKRIALSDKSGKILRNNTONTIDLEQVAGIQRILOQSGSFATASCLICKY 120
DB	309	YPOQFQPSLCHKRIALSDKSGKILRNNTONTIDLEQVAGIQRILOQSGSFATASCLICKY 368
OY	121	KVDCFAVR 128
DB	369	KVDCFAVR 376

```

RESULT      2
09Y6F0      PRELIMINARY:      PRT:      747 AA.
AC 09Y6F0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SIRTUIN TYPE 1.
GN SIRT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA MEDLINE=99310604; PubMed=10381378;
RA Frye R.A.;
RT "Characterization of Five Human cDNAs with Homology to the Yeast SIRT2 Gene; Sirt2-Like Proteins (Sirtulins) Metabolize NAD and May Have Protein ADP-Ribosyltransferase Activity."
RT Biochem. Biophys. Res. Commun. 260:273-279(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Frye R.A.;
RU Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF083106; ADD40849.2; -
DR INTERPRO; IPR003000; -
DR PFAM; PF02146; SIRT2; 1.
SQ SEQUENCE 747 AA; 81680 MW; 2D3BEA6D73DA229F CRC64;

```

```

Query Match      99.7%; Score 680; DB 4; Length 747;
Best Local Similarity 99.2%; Pred. No. 1e-64;
Matches 127; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

YQ 1 IVLTGAGSVSCGIPDRSRDGIYARLAVDPDLPDPOAMFDIYFRKDPFRFAKEI 60
    |||
DB 257 IVLTGAGSVSCGIPDRSRDGIYARLAVDPDLPDPOAMFDIYFRKDPFRFAKEI 316
YQ 61 YPGGFQPSLCHKFTALSDKEGKLRYNTQNTIDLEQVAGIORIQHGSFATASCLICKY 120
    |||
DB 317 YPGGFQPSLCHKFTALSDKEGKLRYNTQNTIDLEQVAGIORIQHGSFATASCLICKY 376
YQ 121 KVDCEAVR 128
    |||
DB 377 KVDCEAVR 384

```

```

RESULT      3
09VK34      PRELIMINARY:      PRT:      823 AA.
AC 09VK34;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SIRT2. PROTEIN.
GN SIRT2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

```

```

RA Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Boulter J., Brockstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RU Science 287:2185-2195(2000).
DR EMBL; AE003639; AAF53248.1; -
DR FLYBASE; FBgn0024291; Sirt2.
DR INTERPRO; IPR003000; -
DR PFAM; PF02146; SIRT2; 1.
SQ SEQUENCE 823 AA; 91836 MW; 3DD1EBA975A263D3 CRC64;

```

```

Query Match      80.6%; Score 550; DB 5; Length 823;
Best Local Similarity 75.8%; Pred. No. 1.e-50;
Matches 97; Conservative 20; Mismatches 11; Indels 0; Gaps 0;

```

```

YQ 1 IVLTGAGSVSCGIPDRSRDGIYARLAVDPDLPDPOAMFDIYFRKDPFRFAKEI 60
    |||
DB 225 IVLTGAGSVSCGIPDRSRDGIYARLAVDPDLPDPOAMFDIYFRKDPFRFAKEI 284
YQ 61 YPGGFQPSLCHKFTALSDKEGKLRYNTQNTIDLEQVAGIORIQHGSFATASCLICKY 120
    |||
DB 285 YPGGFQPSLCHKFTALSDKEGKLRYNTQNTIDLEQVAGIORIQHGSFATASCLICKY 344

```

```

YQ 121 KVDCEAVR 128
    |||
DB 345 KVDCEAVR 352

```

```

RESULT      4
096505      PRELIMINARY:      PRT:      823 AA.
AC 096505;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SIRT2.
GN SIRT2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```


RN [1]
 RP SEQUENCE FROM N.A.
 RA Astrom S U, Rine J.
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF068758; AAC/9684.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A6235E7F CRC64;

Query Match 79.5%; Score 542; DB 5; Length 823;
 Best Local Similarity 75.4%; Pred. No. 7.7e-50;
 Matches 98; Conservative 19; Mismatches 11; Indels 2; Gaps 1;

OY 1 IYLTGAGVSVCIGIPDRSDGIYARLAVDPDLPPQAMFDIEYRRKDPPEFKAKEI 60
 |||||||
 DB 223 IYLTGAGVSVCIGIPDRSTNGIYARLADFPDLPPQAMFDINYPKRDPRPEYKAREI 282
 |||||||
 OY 61 YFG--QFQSLCHKFTALDSKEGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCLIC 118
 |||||||
 DB 283 YGEGRFQFQSPCHRFKMLETKGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCTKC 342
 |||||||
 OY 119 KYKVDCAVR 128
 :::::
 DB 343 RFKNADLR 352

RESULT 5

Q21921 PRELIMINARY; PRT; 607 AA.

AC 021921:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DI R11A8.4 PROTEIN.
 GN R11A8.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cummings P.
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Wellstock L., Wilkinson-Spoat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z70310; CA94364.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 607 AA; 68765 MW; 71C720BABA1DC8BD CRC64;

Query Match 68.2%; Score 465; DB 5; Length 607;
 Best Local Similarity 64.8%; Pred. No. 1e-41;
 Matches 83; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

OY 1 IYLTGAGVSVCIGIPDRSDGIYARLAVDPDLPPQAMFDIEYRRKDPPEFKAKEI 60
 :|||||
 DB 149 IYLTGAGVSVCIGIPDRSDGIYARLAVDPDLPPQAMFDIEYRRKDPPEFKAKEI 208

OY 61 YGEGFQFQSLCHKFTALDSKEGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCLICKY 120
 :|||||
 DB 209 YGEGFQFQSLCHKFTALDSKEGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCLICKY 268
 :|||||
 OY 121 KYDCAVR 128
 :|||||
 DB 269 KYDCAVR 276

RESULT 6

O95889 PRELIMINARY; PRT; 254 AA.

AC 095889:
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DI SIMILAR TO RAT 5E5 ANTIGEN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96207227; PubMed=8619474;
 RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
 RT "A double adaptor method for improved shotgun library
 RT construction."
 RL Anal. Biochem. 236:107-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97264341; PubMed=9110174;
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
 RT "Large-scale concatenation cDNA sequencing."
 RL genome Res. 7:353-358(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wei G., Yu W., Gibbs R.A.;
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF131800; MAD20046.1; -
 DR INTERPRO: IPR003000; -
 DR PFAM: PF02146; SIR2; 1.
 FT NON_TER
 SQ SEQUENCE 254 AA; 28577 MW; 5AEAE01758E476E3 CRC64;

Query Match 48.0%; Score 327.5; DB 4; Length 254;
 Best Local Similarity 50.8%; Pred. No. 2.1e-27;
 Matches 64; Conservative 27; Mismatches 28; Indels 7; Gaps 4;

OY 1 IYLTGAGVSVCIGIPDRSDGIYARLAVDPDLPPQAMFDIEYRRKDPPEFKAKEI 59
 :|||||
 DB 63 IYLTGAGVSVCIGIPDRSDGIYARLAVDPDLPPQAMFDIEYRRKDPPEFKAKEI 120
 :|||||
 OY 60 IYFGQFQFQSLCHKFTALDSKEGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCL- 116
 :|||||
 DB 121 IYFGQFQFQSLCHKFTALDSKEGKLLRNTYNTIDTLEQVAGIQRILQCHGSFATASCL- 180
 :|||||
 OY 117 -ICKY 121
 :|||||
 DB 181 -ICKY 186

RESULT 7
 O9UNT0
 ID O9UNT0: PRELIMINARY; PRT; 352 AA.
 AC O9UNT0:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DI SILENCING INFORMATION REGULATOR 2-LIKE PROTEIN.
 GN SIR2L.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=9931810; PubMed=10393250;
RA Afshar G., Murnane J.P.;
RT "Characterization of a human gene with sequence homology to
RT Saccharomyces cerevisiae SIR2.";
RL Gene 234:161-168(1999).
DR EMBL; AF095714; AAD45871.1; -.
DR INTERPRO; IPR003000; -.
DR PFAM; PF02146; SIR2; 1.
SO SEQUENCE 352 AA; 39515 MW; FFED07DEF9B3416A CRC64;

[illegible]

ID	Q9Y6E9	PRELIMINARY;	PRT;	373 AA.
AC	Q9Y6E9;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)		
DE	SIRTUIN TYPE 2.			
GN	SIRT2			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=TESTIS;			
RC	MEDLINE=99310604; PubMed=10361378;			
RX	Flye R.A.;			
RA	"Characterization of five human cDNAs with homology to the yeast SIRT2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity."			
RT	Biochem. Biophys. Res. Commun. 260:273-279(1999).			
RL	EMBL: AF083107; AAD0850.1; ..			
DR	INTERPRO: IPR003000; ..			
DQ	SEQUENCE 373 AA; 41497 MW; C0985F3D81D5D815 CRC64;			

	Query Match	48.0%;	Score 327.5;	DB 4;	Length 373;
	Best Local Similarity	50.8%;	Pred. No.3.4e-27;		
	Matches	64;	Conservative	27;	Mismatches 28; Indels 7; Gaps 4;
Oy	1	IVLVAGVSVSGCGRDPRS-RDGIYARLAVDPDDLDDQAMDIYFRKDPRPFKAKE	59		
		: : : :			
Dd	64	ICLVAGAGISTAGRIDPFSPTGLVDNI-EKYHLYPEALFEISYFKKHPPFALAKE	121		
Oy	60	IYPGFGPSLCHKFFALDSKEGLLRNTQNIDTLDEOYAGIOR--ILCHGSFAFASCL	116		
	:	: : :			
Dd	122	LYPGFQFTTICHIFERLLKLKGELLCRYQTQNDITLERIARGLEQEDLVNHGTFEYSHCVS	181		

QY	117	-ICKYK	121
		:::	
Db	182	ASCRHE	187

```

RESULT      9
094066      PRELIMINARY;      PRT;      331 AA.
AC      094066;
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE      TRANSCRIPTION REGULATORY PROTEIN.
GN      CA49C4.08C.
OS      Candida albicans (Yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC      anamorphic Saccharomycetales; Candida.
OX      NCBI_TaxID=5476;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1161;
RA      Murphy L., Harris D.;
RL      Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1161;
RA      Barrell B.G., Rajandream M.A.;
RL      Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1161;
RC      MEDLINE=97435544; Pubmed=9290243;
RA      Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT      "A Candida albicans genome project: cosmid contigs, physical mapping
RT      and gene isolation.";
RL      Fungal Genet. Biol. 21:308-314 (1997).
DR      EMBL; AL003503; CAA22018.1; -.
DR      INTERPRO: IPR003000; -.
DR      Pfam: PF02146; SIR2.1.
DR      SEQUENCE      331 AA;      37509 MW;      8748DFB84ED79D10 CRC64;

```

Query Match	46.3%	Score 315.5	DB 3	Length 331
Best Local Similarity	48.8%	Pred. No. 5.7e-26		
Matches	62	Conservative	27	Mismatches 33; Indels 5; Gaps 3
QY	5	GAGSVSCGIDPFRSRD-GIYARLAVDPDLPDQAMEDIEXFRKDPREFFEAKETLYPG	63	
Db	26	GAGISTGAGIDPFRSPDGLVLANLAK--LNLPFAEAVFDIDPFKEDPKFFLYLAELRYG	83	
QY	64	QFQPSLCKHFLALSKEGKLNTYQNDITLLEQVAGIQ--RLQCHGSPATASCLCKTK	121	
Db	84	NEAPRKFHFHFILKLDDGQSLKRVYQNDITLERLAGVEDKYLVEAHGSPASNSCHVDCKHE	143	
QY	122	VDCEAVR	128	
Db	144	MTTEFLK	150	

RESULT	10		
09Y6E8			
09Y6E8	PRELIMINARY;	PRT.	399 AA.
AC	09Y6E8;		
DT	01-NOV-1999 (TrEMBLrel. 12, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)		
DE	SIRTJIN TYPE 3.		
GN	SIRTJ3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_Taxid=9606;		
NN	[1]		

Db 272 КСКНКУД 278

```
Oy      2 VLTGAGVSVSCIGIPDFKRS-RDGIARLAVDPOLPDPQAMFDEIYFKRPPRPFFKA--K 58  
         | : ||| :| :||| | :||| :| :| :| :| :  
Db     34 VMVGAGISVACGIPDFKRPHGTGLAKLS--RYNLNSBEDAFSLPLLRQPVSFYNIILMDM 91
```

QY 59 EIPGQFOPSLCHKFIASDREKLLRNVTNIDTLEQVAGIQR--IIQCHSFFATASCL 116
 Db 92 DLMFGKCYCPTVYHNFHISILAKKGMILCCCTONIDLERACGIPESLVEAHGHSFSSACV 151
 QY 117 ICKYVDCENAR 128
 Db 152 DCHAKYDINIAR 163

RESULT 14
 Q9WYWO PRELIMINARY; PRT; 246 AA.
 AC Q9WYWO;
 DT 01-NOV-1999 (TREMBLER, 12, Created)
 DT 01-NOV-1999 (TREMBLER, 12, Last sequence update)
 DT 01-JUN-2000 (TREMBLER, 14, Last annotation update)
 DE REGULATORY PROTEIN, SIR2 FAMILY.
 GN TM0490.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 Stewart A.M., Cotton M.D., Pratt W.S., Phillips C.A., Richardson D.,
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1998).
 DR EMBL: AE001726; AAD35575.1;
 DR TIGR: TM0490;
 DR INTERPRO: IPR003000;
 SQ SEQUENCE 246 AA; 27538 MW; 6BD2EF1B751C190D CRC64;

Query Match 33.8%; Score 230.5; DB 2; Length 246;
 Best Local Similarity 38.8%; Pred. No. 5.5e-17;
 Matches 50; Conservative 26; Mismatches 42; Indels 11; Gaps 3;

QY 1 IVLTGAGVSCGIPDFSRDGIYARLAVDPDLPDQAMFDIEYFRKDPPEFFAKE- 59
 Db 17 VLTGAGSTFSGIPDFSGPNGIYKYS-----ONVFDIDFFYSHPEFFYFAKEG 67
 QY 60 IYPG-QFQPSLCHKFIASDREKLLRNVTNIDTLEQVAGIQRILQCHSFFATASCLIC 118
 Db 68 IFPMQAKRPNLAHVLAATLEKEGLIEAVITQNDIRLHQKASKVYELHGVNVEEYCYRC 127
 QY 119 KYKVDCEAV 127
 Db 128 EKKYTVEDV 136

RESULT 15
 Q07595 PRELIMINARY; PRT; 247 AA.
 AC Q07595;
 DT 01-JUL-1997 (TREMBLER, 04, Created)
 DT 01-JAN-1998 (TREMBLER, 05, Last sequence update)
 DT 01-OCT-2000 (TREMBLER, 15, Last annotation update)
 DE HYPOTHEMETICAL 27.4 KDA PROTEIN.
 GN YH02.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RA Nuback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=94156839; PubMed=8113174;
 RA Huilet F.M., Lee J., Shi L., Sun G., Chesnut R., Sharkova E.,
 RA Duggan M.F., Kapp N.;
 RT "Sequential action of two-component genetic switches regulates the PHO
 RT regulon in Bacillus subtilis."
 RL J. Bacteriol. 176:1348-1358(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=94321341; PubMed=8045899;
 RA Jin S., Sonenshein A.L.;
 RT "Transcriptional regulation of Bacillus subtilis citrate synthase
 RT genes."
 RL J. Bacteriol. 176:4680-4690(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolyard A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haeich J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nuback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogikawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
 RA Priesean E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassartot A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y14082; CA74510.1;
 DR EMBL: Z99109; CAB12804.1;
 DR INTERPRO: IPR003000;
 DR PFM: PFO2146; SIR2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 247 AA; 27417 MW; 8BAD6F797D3A99F7 CRC64;

Query Match 32.0%; Score 218.5; DB 2; Length 247;
 Best Local Similarity 39.4%; Pred. No. 1.1e-15;
 Matches 50; Conservative 22; Mismatches 44; Indels 11; Gaps 3;

QY 1 IVLTGAGVSCGIPDFSRDGIYARLAVDPDLPDQAMFDIEYFRKDPPEFFAKEI 60

```

Db      15  VVLTGAGMSTESGIPDPERRSAGIWT-----DASRMEAM-SLDYFLSYRLEFMPKKEL 67
QY      61  Y---PGOFOPSLCHKFIALSDKEGKLLRNYTONIDTLEOVAGIORILOCHGSPATASCL 116
Db      68  FOMKMSGSPFEPNEGHLLAELERKOGKQVDIFTONIDGLHKKAGSRHYVELHGSIGTACP 127
QY      117  ICKYKVD 123
Db      128  ACGARYD 134

```

Search completed: February 16, 2001, 10:51:35
 Job time: 595 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:51:35 ; Search time 137.23 Seconds
(without alignments)
279.290 Million cell updates/sec

Title: US-09-461-580A-12
Perfect score: 1732
Sequence: 1 VINILSEPPKRRKKRDINTL.....PLHISEDSSSPERTVPDSS 327

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_15:*
2: sp.archaea:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1732	100.0	737	11	09QXG8	09QXG8 mus musculu
2	1699	98.1	747	4	09YGF0	09YGF0 homo sapien
3	1021	58.9	823	5	09VK34	09VK34 drosophila
4	1009.5	58.3	823	5	096505	096505 drosophila
5	847	48.9	607	5	021921	021921 caenorhabdi
6	575.5	33.2	471	3	094640	094640 schizosacch
7	555.5	32.1	352	4	09UNTO	09UNTO homo sapien
8	522.5	32.1	373	4	09YCE9	09YCE9 homo sapien
9	522.5	30.2	331	4	094066	094066 candida alb
10	508	29.3	399	3	09YGE8	09YGE8 homo sapien
11	504.5	29.1	332	3	09USN7	09USN7 schizosacch
12	492	28.4	254	4	095889	095889 homo sapien
13	406	23.4	351	5	096670	096670 trypanosoma
14	404.5	23.4	373	4	09PGC8	09PGC8 homo sapien
15	349	20.2	247	2	007595	007595 bacillus su
16	346	20.0	246	2	09WY20	09WY20 thermotoga
17	330.5	19.1	253	1	030124	030124 archaeoglob
18	295	17.0	245	1	028597	028597 archaeoglob
19	285.5	16.5	247	1	09YB13	09YB13 aeropyrum p

20	282.5	16.3	249	1	058669	058669 pyrococcus
21	267.5	15.4	250	1	09UZ87	09UZ87 pyrococcus
22	264	15.2	243	2	053700	053700 staphylococ
23	262	15.1	310	4	09YCE6	09YCE6 homo sapien
24	251.5	14.5	299	4	09NKA8	09NKA8 homo sapien
25	251	14.5	415	3	09UR39	09UR39 schizosacch
26	245	14.1	239	2	067919	067919 aquifex aeo
27	237	13.7	314	4	09YGE7	09YGE7 homo sapien
28	233	13.5	326	4	043346	043346 homo sapien
29	230	13.3	257	2	09KEE5	09KEE5 bacillus ha
30	220	12.7	287	5	020480	020480 caenorhabdi
31	216.5	12.5	299	2	09RL35	09RL35 streptomyce
32	211	12.2	237	2	006549	006549 mycobacteri
33	210.5	12.2	377	4	09UND1	09UND1 homo sapien
34	210	12.1	259	2	09KRX4	09KRX4 vibrio chol
35	210	12.1	279	2	09Y960	09Y960 escherichia
36	206.5	11.9	355	4	09NRC7	09NRC7 homo sapien
37	199.5	11.5	208	2	09ZAB8	09ZAB8 actinobacil
38	197.5	11.4	234	2	09ZJW8	09ZJW8 helicobacte
39	190	11.0	287	5	020481	020481 caenorhabdi
40	187.5	10.8	233	2	09PNP1	09PNP1 campylobact
41	186	10.7	227	2	09JN05	09JN05 campylobact
42	185.5	10.7	246	2	09RYD4	09RYD4 delnococtus
43	185	10.7	320	5	09UIA6	09UIA6 leishmania
44	183	10.6	483	10	09XGTL	09XGTL oryza sativ
45	180.5	10.4	204	5	09W4A1	09W4A1 drosophila

ALIGNMENTS

RESULT	1	ALIGNMENTS
09QXG8	1	
ID	09QXG8	PRELIMINARY; PRT; 737 AA.
AC	09QXG8:	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE	SIR2ALPHA PROTEIN.	
GN	SIR2ALPHA.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RP	STRAIN=SWISS WEBSTER/NIH;	
RA	Imai S., Armstrong C.M., Guarente L.;	
RT	"Silencing and aging protein Sir2 is an NAD-dependent histone	
RT	deacetylase.";	
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF214646; AAF24983.1; -	
DR	HSSP; P04002; IATE.	
DR	INTERPRO; IPR003000; -	
PR	PRAM; PR02146; SIR2; 1.	
SQ	SEQUENCE 737 AA; 80371 MW; 7F15625E29433119 CRC64;	
Query Match	100.0%; Score 1732; DB 11; Length 737;	
Best Local Similarity	100.0%; Pred. No. 1.6e-138;	
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 VINILSEPPKRRKKRDINTEDAVKLLQECKRIIVLTGAGVSSCGIPDFSRDGIYARL 60	058669 pyrococcus
DB	216 VINILSEPPKRRKKRDINTEDAVKLLQECKRIIVLTGAGVSSCGIPDFSRDGIYARL 275	09UZ87 pyrococcus
QY	61 AVDPDPLDPQAMDIYFRKDPFRPFKAKEIYPGQFQPSLCKHFTALSSEKELLRNY 120	053700 staphylococ
DB	276 AVDPDPLDPQAMDIYFRKDPFRPFKAKEIYPGQFQPSLCKHFTALSSEKELLRNY 335	09YCE6 homo sapien
QY	121 TONIDTLEQVAGIORIILQCHGSFATASCLCKYKVDCAVAGDIFFNQVVPKPCPADPEP 180	09NKA8 homo sapien
DB	336 TONIDTLEQVAGIORIILQCHGSFATASCLCKYKVDCAVAGDIFFNQVVPKPCPADPEP 395	09UR39 schizosacch

```

QY 181 LAIMKPELVFEGENLPEQFHRAMKYDKDEVDLLIVTGSLSKVRPALPSSIPHEVPOIL 240
DB 396 LAIMKPELVFEGENLPEQFHRAMKYDKDEVDLLIVTGSLSKVRPALPSSIPHEVPOIL 455
QY 241 INREPLPHLHFDVLLGDCDVIINELCHRLGGEVAKLCCNPVKLSITEKPPROKELVH 300
DB 456 INREPLPHLHFDVLLGDCDVIINELCHRLGGEVAKLCCNPVKLSITEKPPROKELVH 515
QY 301 LSELPPPLHISEDSSSPERTVPODSS 327
DB 516 LSELPPPLHISEDSSSPERTVPODSS 542

RESULT 2
QY96F0 PRELIMINARY: PRT: 747 AA.
AC QY96F0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SIRTUIN TYPE 1.
GN SIRT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=99310604; PubMed=10381378;
RA Fyfe R.A.;
RT "Characterization of Five Human cDNAs with Homology to the Yeast SIRT2 Gene; Sirt2-like Proteins (Sirtuins) Metabolize NAD and May Have
RT Protein ADP-Ribosyltransferase Activity.";
RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Fyfe R.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083106; AAD0849.2; -
DR INTERPRO; IPR003000; -
DR PFM; PF02146; SIRT2; 1.
SQ SEQUENCE 747 AA; 81680 MM; 2D3BE6D73DA229F CRC64;

Query Match 98.1%; Score 1699; DB 4; Length 747;
Best Local Similarity 97.9%; Pred. No. 1e-135;
Matches 320; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

DB 524 LSELPPPLHISEDSSSPERTSPDSS 550

RESULT 3
QY9K34 PRELIMINARY: PRT: 823 AA.
ID QY9K34;
AC QY9K34;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SIRT2 PROTEIN.
GN SIRT2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Baldwin D.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Basley E.M.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brinkstein P., Brothier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Ibovym C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svyrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Rubin J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003639; AAF53248.1; -
DR FLYBASE; FBgn0024291; SIRT2.
DR INTERPRO; IPR003000; -
DR PFM; PF02146; SIRT2; 1.
SQ SEQUENCE 823 AA; 91836 MM; 3DD1EBA975A263D3 CRC64;

Query Match 58.9%; Score 1021; DB 5; Length 823;
Best Local Similarity 58.7%; Pred. No. 3.2e-76;
Matches 203; Conservative 46; Mismatches 71; Indels 26; Gaps 5;

```



```

QY 63 DFDLPDPOAMFIDIEYRKDPREFKFAKIEYQFOPSLCHKEFKIALSDKEGKLLRNYTQ 122
DB 254 DFDLPDPOAMFIDIEYRKDPREFKFAKIEYQFOPSLCHKEFKIALSDKEGKLLRNYTQ 313
QY 123 NIDTLEQVAGIORTLOCHGSFATASCLICKYKVCCEAVRDIFNOVPRCP- 176
DB 314 NIDTLEQVAGIORTLOCHGSFATASCLICKYKVCCEAVRDIFNOVPRCP- 373
QY 177 -----ADEPL-----AIMKPEIVFEGENLEPQFHRAMKYDKDEVLLIVIGSSLKVR 223
DB 374 VDAVAATEBELQOLVENGIMKPDIVFEGEGELPDEYHTVATDKVCDLLIVIGSSLKVR 433
QY 224 VPAVLISSIPHEVPOILINREPLPHLHFDVELLIGDGVIIINELCHRLGGE---YAKLCCN 280
DB 434 VPAVLISSIPHEVPOILINREPLPHLHFDVELLIGDGVIIINELCHRLSDNDCCWROLCCD 493
QY 281 PVKLISETTEKPPROKELVHSELPTPLHISEDSSSPERTVPODS 326
DB 494 ESVLTESKELMP-PEHSHNHLH---HLHHRHCSSSEKQSOQLDT 535

RESULT 4
ID 096505 PRELIMINARY; PRT; 823 AA.
AC 096505:
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE SIR2.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Astrom S.U., Rine J.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF068758; AAC79684.1; -
DR INTERPRO: IPR003000; -
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A46235E7F CRC64;

Query Match 58.3%; Score 1009.5; DB 5; Length 823;
Best Local Similarity 59.3%; Pred. No. 3e-77;
Matches 200; Conservative 45; Mismatches 67; Indels 25; Gaps 5;

```

```

RESULT 5
ID 021921 PRELIMINARY; PRT; 607 AA.
AC 021921:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE RIL8.4 PROTEIN.
GN RIL8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cummings P.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL: Z70310; CAA94364.1; -
DR INTERPRO: IPR003000; -
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 607 AA; 68765 MW; 71C720BABA4DC8BD CRC64;

Query Match 48.9%; Score 847; DB 5; Length 607;
Best Local Similarity 49.3%; Pred. No. 1.2e-63;
Matches 168; Conservative 60; Mismatches 87; Indels 26; Gaps 5;

```


Qy	62	VFDPFLPPOAMFIEKFKRQPRPEFFAKFITYIGQSPISCHHFIASDKEGKLKNT	121
Db	92	-EKHLLPPEAIFETSTFKKREFFPALKELVPGQKRPITCIHFMYLKDGLKLCYT	150
Qy	122	QNIDTLEOVAGIOR--IIQCHGSFATASCL--ICKYVVDENAVGDIENFQVPPCPRCPA	177
Db	151	QNIDTLERLACLEDLEDVLEANGFTYTSICVSAASCRHEYP LSWMKKEKLFSEVTPCEDCQ-	209
Qy	178	DEPLAIKPELVFFGGENLPEDGFHRAKRYDKDEVDLLIVIGSSSLKRVPAVALIPSIPIHEVP	237
Db	210	---SLVPPDIVFEGESLPARFFSCMOSEDFLKVDLLVMGTSILOVQPFASILCKAPLSTP	265
Qy	238	QILINREPL---PDL-----HFVDLLGDCDVIINELCHRLG-----	271
Db	266	RLINKERAGSDPFLGIMINGCGMDPKSKAYRDVAWMLGECQGLALAEELGWMKEL	323
Qy	272	-----GEYAKLCC-----NPVRLSEITEKPPRPQKRELVLHSELPLPLHISEDSSSPE	319
Db	326	EDLVREHESIDAGSGAGVNP-STSAAPKSPPPAK-----DEANTTE	368
Qy	320	RTVPQ 324	
Db	369	REKPPQ 373	
RESULT	9		
Q04066			
ID	094066	PRELIMINARY:	PRT: 331 AA.
AC	094066;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	TRANSCRIPTION REGULATORY PROTEIN.		
GN	CA99C4.08C.		
OS	Candida albicans (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;		
OC	anamorphic Saccharomycetales; Candida.		
OX	NCBI_TaxId=5476;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-1161;		
RC	STRAIN-1161;		
RA	Barrell B.G., Rajandream M.A.;		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-1161;		
RC	STRAIN-1161;		
RA	Tate E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;		
RT	"A Candida albicans genome project: cosmid contigs, physical mapping,		
RT	and gene isolation.";		
RL	Fungal Genet. Biol. 21:308-314 (1997).		
DR	EMBL: AL033503; CA22018.1; -		
DR	INTERPRO: IPR003000; -		
PR	PF02146; SIR2.1;		
Q0	SEQUENCE 331 AA; 37509 MW; 8748DFB44ED79D10 CRC64;		

[illegible]

Db	61	FDIDFKDKRPKRYTLAEELYPGNFAPYKPHHFIKLLQDQSLKRVYQTQIDTLERLAGV	120
Qy	134	Q--RLQCHGSFATASCLICKYKVDCEAVRGDIDNQVWRPCRPRADEPLAIKPELVET	192
Db	121	EDKYIVHGHSFASPNHCDCDCKHEKMTTEFLTKYMDKKIPSCQHE-----GYVKPDIIVF	175
Qy	192	GENIPQOPHAKMKYDKDQVDLLIYIGSSIAKYRPVALIPSSIPHEVPQILINREPLPLHF	252
Db	176	GGGLPWFKEFDIMEDDCDEVEVAIYAGTSLVTFPPASLPGEVYKCKLVRVAKVEGVTFKH	235
Qy	252	-----DVELLDCDVIITINELCHRIAGSEYAKLCKNPVLKSEITEKPRPPOKELVHLSLPP	306
Db	236	EPKRSIDIALLHDCDQIVAKELCTLGLD-----DKLNEYER-----EKIKYSKATTKK	263
Qy	307	TPPHISEDSSSPERTVPD 325	
Db	284	TKMHEIDEDLKEEAHLKED 302	
RESULT	10		
Qy9Y6E8			
ID	Q9Y6E8	PRELIMINARY;	PRT; 399 AA.
AC	Q9Y6E8;		
DT	01-NOV-1999 (TReMBLrel_12, Created)		
DT	01-NOV-1999 (TReMBLrel_12, Last sequence update)		
DT	01-JUN-2000 (TReMBLrel_14, Last annotation update)		
DE	SIRTUIN TYPE 3.		
GN	SIRT3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TESTIS;		
RX	MEDLINE=99310604; PubMed=10381378;		
RA	Frye R.A.;		
RT	"Characterization of five human cDNAs with homology to the yeast Sir2		
RT	gene: Sir2-like proteins (sirtuins) metabolize NAD and may have		
RT	protein ADP-ribosyltransferase activity.";		
RL	Biochem. Biophys. Res. Commun. 260:273-279(1999).		
DR	EMBL: AF083108; AAD40851.1; .		
DR	INTERPRO: IPR003000; .		
SO	SEQUENCE 399 AA; 43573 MW; 4BABBD3AC5FC7901 CRC64;		

[illegible]

ID	ORGUSN7	PRELIMINARY:	PRT:	332 AA.
AC	09USN7.			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-OCT-2000 (TREMBlrel. 15, Last annotation update)			
DE	PUTATIVE SIR2-LIKE TRANSCRIPTIONAL REGULATORY PROTEIN.			
GN	SPC132.02.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;			
OC	Schizosaccharomycetaceae; Schizosaccharomyces.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972H-;			
RA	Lyne M., Rajandream M.A., Barrell B.G., Aert R., Volckaert G.;			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL121807; CAB58129.1; -.			
DR	INTERPRO: IPR003000; -.			
DR	PFAM: PF02146; SIR2; 1.			
SO	SEQUENCE	332 AA;	37915 MW;	936A08D240C8BCFF CRC64;

Query Match	29.1%;	Score 504.5;	DB 3;	Length 332;
Best Local Similarity	40.6%;	Pred. No. 6e-35;		
Matches 114;	Conservative 53;	Mismatches 95;	Indels 19;	Gaps 8;

QY	20	I EDAVKLLQEE--CKKIIIVLTGAGVSVCIGIPPEFRSND--GIYARLAVDPDLPDDPMPEI	76
Db	15	LEKVASLIKKEGVKKKICVWAGVGISTAGCIPPFRRSPETIYNNL--QRNPLTAAVEPDL	72
QY	77	EYFRKDPPPEFKAEIYVPGQFQPSLCKHKEIALSDKEGKLNNYTONIDITLQOVAGI--Q	134
Db	73	SYFRKNPPPEVYELAHMELMEKRYPRPYTHYFIRLTHDKRLQCYCTONIDITLERLAGVPK	132
QY	135	RILQGHSEFATASCLICKKYKVDCEAVRGDIFNQVYPRCGRCADEPLAMKEIYVFGGN	194
Db	133	ALTEHGSGFOYSRCIECEYEMATEYEVKRCITMOKVPRKSCSK-----GIKIMPIYVGGG	187
QY	195	LPEQFRHAKMYDKDEVDLLIYVSSLLKRPVALLPSSIPHEVPOQLINREP--LPHTLF	251
Db	188	LPMPFEHEMKQTKVCDMALVIGTSLVHPFADPLDEIYVANKQRYVLINPEAGDGEGRKK	247
QY	232	DVELLGGDCDVIITNELCHRLG--GEFAKLCCNRYKALSETLEK	290
Db	248	DIMILGDCDSQVRLCKLLGWSDELEKRLDYSVE--TLTEE	286

RESULT	12			
095889				
ID	095889	PRELIMINARY;	PRT;	254 AA.
AC	095889;			
DT	01-May-1999	(TREMBLrel. 10, created)		
DT	01-May-1999	(TREMBLrel. 10, last sequence update)		
DT	01-Oct-2000	(TREMBLrel. 15, last annotation update)		
DE	SIMILAR TO RAT 5E5 ANTIGEN (FRAGMENT).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96207227; PubMed=8619474;			
RA	Anderson B., Wentland M.A., Ricefrente J.Y., Liu W., Gibbs R.A.;			
RT	"A 'double adaptor' method for improved shotgun library			
RT	construction.";			
RL	Anal. Biochem. 236:107-113(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97264341; PubMed=9110174;			
RA	Yu W., Anderson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,			
RA	Ricefrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;			
RT	"Large-scale concatenation cDNA sequencing.";			
RL	Genome Res. 7:353-358(1997).			

RN [3]
 RP SEQUENCE FROM N.A.
 RA Mel G. Yu M., Gibbs R.A.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL, AF131800; NAD20046.1; -.
 DR INTERPRO; IPR003000; -.
 DR PFM; PF02146; STR2; 1.
 FT NON_TER 1
 SQ SEQUENCE 254 AA; 28577 MW; 5AEAD0158EA76E3 CRC64;

Query Match	28.48;	Score 492;	DB 4;	Length 254;
Best Local Similarity	43.78;	Pred. NO. 4.9e-34;		
Matches 100; Conservative	55;	Mismatches 60;	Indels 14;	Gaps 6;

Qy	5	LSPEPRKKRRKDINTIEDAVKLLQ--EKKIITVLTGAUVSSCIGPDRDS-RDGIYAPALA	61
Db	32	LSLGSQKEKLDELLEGVARYMOSEKRRRVATCLGAGIAISTAGIPDRDSSTGTGLYNL-	90
Qy	62	VDFPQLPDPQAMFIDIEYFRKDRPPEFKAEKELYPCQGFQPSLCHKHFTALSDREGKLLRNYT	121
Db	91	-EKYLLPPEAFELFESTYKKKHPEPFALAKELYPQGFQFTYICHYMRLLKDKGILLRQYT	149
Qy	122	QNIDTLEQVAGIQR--ILQCHGSFATASCL--ICKYKVDCEAVRGDIENQVAPRCPRCPA	177
Db	150	QNIDTLERLAGDEQDILVEAHGTFYTSHCVSASCSAREKPLSMWKKEKIFSEVTPKCEDCO-	208
Qy	178	DEPLAIKMPETLYFEGENLPPEQPHRAMKRDKDEVDLLIYIGSSLAKYRPA	226.
Db	209	---SLVMPDVIYFGESELPAREFSCQMSQDFLKVLLDLVLLVYSLQGRGIA	253

RESULT 13		
ID	096670	PRELIMINARY; PRT; 351 AA.
AC	096670;	
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE	SIR2 HOMOLOG.	
GN	SIR2.	
OS	Trypanosoma brucei.	
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.	
OX	NCBI_TaxID=5691;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=427;	
RA	Hoek M., Wirtz E., Cross G.A.M.;	
RT	"Cloning of a Trypanosoma brucei Sir2 homolog.";	
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF102869; AAC73004.1; -.	
DR	INTERPRO: IPR003000; -.	
DR	PFAM; PF02146; SIR2; 1.	
SQ	SEQUENCE 351 AA; 38529 MW; 9EA0BD5106670575 CRC64;	

[illegible]

```

Db 178 G-----GIKPPVVFEGENLPEAFNVAGL-IBETELLITLSTQVHPADALAMVPSD 231
QY 236 VQIILNREPPLH-----251
Db 232 VQVLENLGRMRFPDRTPNFRASSYRLSTGNGSKISSDSSSSVDGYDOP 291
QY 252 -----DVELLGDGVITINELCHRLG-GEYAKLCCNPVKLSEITEKPPPOK 296
Db 292 TLAENDETGVLRDIFFFPGDCQSVSRFAQLGFE--QLDASVREGREIFERTIRREK 347

RESULT 14
09P068 PRELIMINARY: PRT: 373 AA.
AC 09P068:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE SILENT INFORMATION REGULATOR 2 HOMOLOG.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADRENAL GLAND;
RA Ren S., Fu S., Gu J., Jin W., Gu Y., Huang Q., Dong H., Yu Y., Fu G.,
RA Wang Y., Chen Z., Han Z.;
RA "A novel gene expressed in human adrenal gland.";
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC TISSUE-ADRENAL GLAND;
RA Peng Y., Gu Y., Li Y., Fu S., Gu J., Zhang L., Jiang C., Yu Y.,
RA Han Z., Wang Y., Chen Z., Fu G.;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF160214; AAF67015.1;
SQ SEQUENCE 373 AA; 40737 MW; 81EAA55453F3638F CRC64;

Query Match 23.4%; Score 404.5; DB 4; Length 373;
Best local Similarity 30.3%; Pred. No. 2,1e-26;
Matches 114; Conservative 62; Mismatches 109; Indels 91; Gaps 14;

QY 5 LSEPPRRKKRKINTEDAVKLIQ--ECKKIYLTGAGSVSCGIDPERS-RDGIYARLA 61
Db 33 LSLGSKERLDELTEGVARYMQSEKRCRVICLVAGISTAGIPDRSPGLVDNL- 91
QY 62 VDEPDLDDQAMEDIYFRKDRPFKAKIYPOGO-----PSLC- 103
Db 92 -EKYHLPYPALEISYFKHPPFPALAKELPGQFQANHSLAHAPAGCATPALH 150
QY 104 --HKFTALSDKEKGLRNYQNTIDTEOVAGIORIILQCHSPATASCLIKYKDCAVR 161
Db 151 AEHRYPGANSRAG-----TGLGGGAGHLL--HTLRHAN--CRHEYPISWMK 194
QY 162 GDIFNQVPRCPADDEPLAKKPEIVFEGENLPEOFHRAKRYDKDEVLLIVIGSSLK 221
Db 195 EKIFSEVTPTCEDCQ-----SLVKPDIVFEGESLPAFFSCMSQDPLKVLVLMGTSIQ 249
QY 222 VRIVALIPSSIPHEVPOIILNREPL-----PHL-----HFDVELLGDCCD 260
Db 250 VQFPAFLISAPLSTPLLNKKEKAGOSDFPLMIMGLGGMDFSKKAYRDVAMGECDD 309
QY 261 VIINELCHRLGGEYAKLCCNPVKLSEITEK-----PPRPOKELVHLSLEPTP 308
Db 310 QGGLALAEELLGK-----KELEDLVPRGARGHRCPRVGGGPNPSTASPKKSPRA- 360
QY 309 LHISESSSPERTVPQ 324
Db 361 ---KDEARTTEREKPO 373

```

```

RESULT 15
ID 007595 PRELIMINARY: PRT: 247 AA.
AC 007595:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE SILENT INFORMATION REGULATOR 2 HOMOLOG.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADRENAL GLAND;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94156839; PubMed=8113174;
RA Hulet F.M., Lee J., Shi L., Sun G., Chesnut R., Sharkova E.,
RA Duggan M.F., Kapp N.;
RT "Sequential action of two-component genetic switches regulates the PRO
RT regulon in Bacillus subtilis.";
RL J. Bacteriol. 176:1348-1358(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94321341; PubMed=8045899;
RA Jin S., Sonenshein A.L.;
RT "Transcriptional regulation of Bacillus subtilis citrate synthase
RT genes.";
RL J. Bacteriol. 176:4680-4690(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Brun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Eutian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudega B., Park S.H.,
RA Paricio V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiuchi J., Sekowska A., Serot S.J., Serot P., Shin B.S., Soldo B.,
RA Sorokin A., Tacom E., Takagi T., Takahashi H., Takemaru K.,
RA Takenuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE FROM N.A.

```

RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
RL Submitted (NOV-1997) to the EMBL/genbank/DBJ databases
DR EMBL: Y14083; CAA14510.1; -
DR EMBL: 299109; CAA12804.1; -
DR INTERPRO: IPR003000;
DR Pfam; PF02146; S1R2; 1.
KW Hypothetical protein.
SQ SEQUENCE 247 AA; 27417 MW; 8BAD6F797D3A99F7 CRC64;

Query Match	20.28;	Score	349;	DB	2;	Length	247;
Best Local Similarity	32.78;	Pred. No.	6.2e-22;				
Matches	87;	Conservative	54;	Mismatches	91;	Indels	34;
						Gaps	9;

```

QY      20  IEDAVKLIOEKKKIIIVLTGAVSVSCGIPDRSRSDGIYARLAVDFPLPOLPOAAMFDEYE  79
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      1  METEKSLIHEORIVLTLTGAGMSTESGIPDRSGAGIYTE-----DASRMEAM-SLDYF  53

QY      80  RKDPRPEFKAEIY---PQGFQPSLCHMFIALDSKGLLRNYTONIDTLEOVAGIQR  135R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      54  LSYRLFLPKRKELFOMKMSGSEFPENEGHLLTIAELERQGVDFITONIGLHKHKSRRH  113R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      136  ILOHGSATATSLCLICKKVKVCEAVRDIQNOVVPRQPCRADEPL--ALMKPEIYEGE  193R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      114  VYELHGSIQTLACRCACGAKRDLP-----HLEREVPETAAAGNMDIGCYLAKTDIVLFGD  168R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      194  NLPEQFIRAMKYDK-DEYDILLIVIGSSLIKARVALID--SIPHEVQILINREPLHLH  250R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      170  AV---MHEDPLYEKDAQDILLVLTGTSLEVARFARVEDASLIPIGMKVIINLEPT----  222R
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      251  FDELVLGCDVYIINELCHRLGGEYAK  276
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      223  -----YCDSLFMDVYIHKQIGGEFAR  241
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Search completed: February 16, 2001, 10:51:37
Job time: 597 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 16, 2001, 10:51:37 ; Search time 137.23 Seconds
(without alignments)
90.534 Million cell updates/sec

Title: US-09-461-580A-19
Perfect score: 556
Sequence: 1 IIVLTGAGVSGVSCGIPDRS.....NYTONIDTLEQVAGIQRILQ 106

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_15:*
2: sp-archaea:*
3: sp-bacteria:*
4: sp-fungi:*
5: sp-human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp-organella:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	556	100.0	737	11	09QXG8	09QXG8 mus musculu
2	554	99.6	747	4	09Y6F0	09Y6F0 homo sapien
3	471	84.7	823	5	09YK34	09YK34 drosophila
4	463	83.3	823	5	096505	096505 drosophila
5	410	73.7	607	5	021921	021921 caenorhabdi
6	301.5	54.2	254	4	095889	095889 homo sapien
7	301.5	54.2	352	4	09UNTO	09UNTO homo sapien
8	301.5	54.2	373	4	09Y6E9	09Y6E9 homo sapien
9	275.5	49.6	331	3	094066	094066 candida alb
10	263.5	47.4	399	4	09Y6E8	09Y6E8 homo sapien
11	243.5	43.8	332	3	09USN7	09USN7 schizosacch
12	226	40.6	471	3	094640	094640 schizosacch
13	199.5	35.9	246	2	09YXW0	09YXW0 thermotoga
14	196	35.3	373	4	09Y6G8	09Y6G8 homo sapien
15	194.5	35.0	351	5	096670	096670 trypanosoma
16	179.5	32.3	247	2	007595	007595 bacillus su
17	174.5	31.4	253	1	030124	030124 archaeoglob
18	162.5	29.2	415	3	09UR39	09UR39 schizosacch
19	153	27.5	243	2	053700	053700 staphylococ

20	153	27.5	247	1	09YB13	09YB13 aeropyrum p
21	149	26.8	245	1	028597	028597 archaeoglob
22	138	24.8	239	2	067919	067919 aquifex aeo
23	132	23.7	249	1	058669	058669 pyrococcus
24	129.5	23.3	259	2	09KRX4	09KRX4 vibrio chol
25	129	23.2	299	4	09NXX8	09NXX8 homo sapien
26	129	23.2	310	4	09Y6E6	09Y6E6 homo sapien
27	127.5	22.9	279	2	P75960	P75960 escherichia
28	127.5	22.9	314	4	09Y6E7	09Y6E7 homo sapien
29	127.5	22.9	326	4	043346	043346 homo sapien
30	123	22.1	299	2	09RL35	09RL35 streptomyce
31	120.5	21.7	233	2	09PNP1	09PNP1 campylobact
32	119	21.4	246	2	09RYD4	09RYD4 delnoccocus
33	119	21.4	250	1	09UZE7	09UZE7 pyrococcus
34	118.5	21.3	204	5	09W4A1	09W4A1 drosophila
35	118.5	21.3	287	5	020480	020480 caenorhabdi
36	115.5	20.8	208	2	09ZAB8	09ZAB8 actinobacti
37	115.5	20.8	234	2	09ZJW8	09ZJW8 helicobacte
38	113	20.3	237	2	09KEE5	09KEE5 bacillus ha
39	110.5	19.9	287	5	020481	020481 caenorhabdi
40	108.5	19.5	227	2	09JN05	09JN05 campylobact
41	105	18.9	483	10	09XGT1	09XGT1 oryza sativ
42	103	18.5	87	2	09ZB11	09ZB11 streptomyce
43	102	18.3	91	5	09NMB2	09NMB2 leishmania
44	102	18.3	1095	5	001478	001478 caenorhabdi
45	100.5	18.1	317	5	09VH08	09VH08 drosophila

ALIGNMENTS

RESULT 1
ID 09QXG8 PRELIMINARY; PRT; 737 AA.
AC 09QXG8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE STR2ALPHA PROTEIN.
GN STR2ALPHA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIH;
RA Imai S., Armstrong C.M., Guarente L.;
RT Silencing and aging protein Sir2 is an NAD-dependent histone
RT deacetylase.*;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF214646; AAF24983.1; -
DR HSSP; P04002; IATF.
DR INTERPRO; IPR003000; -
DR PRAM; PR02146; SIR2; 1.
SQ SEQUENCE 737 AA; 80371 MW; 7F15625E29433119 CRC64;

Query Match 100.0%; Score 556; DB 11; Length 737;
Best Local Similarity 100.0%; Pred. No. 6.9e-53;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSGVSCGIPDRSDGIYARLAVDPDLPQAMFDIEYFRKDPPEFFAKE 60
DB 248 IIVLTGAGVSGVSCGIPDRSDGIYARLAVDPDLPQAMFDIEYFRKDPPEFFAKE 307
QY 61 IIVGQFQPSLCHKFIALSDKEGKILRYTONIDTLEQVAGIQRILQ 106
DB 308 IIVGQFQPSLCHKFIALSDKEGKILRYTONIDTLEQVAGIQRILQ 353
RESULT 2
QY6F0

ID Q9Y6F0 PRELIMINARY; PRT: 747 AA.
 AC Q9Y6F0;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE SIRTUIN TYPE 1.
 GN SIRT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=99310604; PubMed=10381378;
 RA Fyfe R.A.;
 RT "Characterization of Five Human CDNs with Homology to the Yeast SIRT
 Gene; Sir2-Like Proteins (SirTuins) Metabolize NAD and May Have
 Protein ADP-Ribosyltransferase Activity.";
 RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Fyfe R.A.;
 RL Submitted (Apr-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AF083106; AAD40849.2; -
 DR INTERPRO: IPR003000; -
 DR PFM: PFM02146; SIR2; 1.
 SQ SEQUENCE .747 AA; 81680 MW; 2D3BEA6D73DA229F CRC64;

Query Match 99.6%; Score 554; DB 4; Length 747;
 Best Local Similarity 99.1%; Pred. No. 1.2e-52;
 Matches 105; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 IIVLTGAGVSVCIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
 |||||||
 Db 256 IIVLTGAGVSVCIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 315
 |||||||
 Oy 61 IYPGQFQPSLCHKFTALSDKKGKLLRNTONTIDLEQVAGIORTIIO 106
 |||||||
 Db 316 IYPGQFQPSLCHKFTALSDKKGKLLRNTONTIDLEQVAGIORTIIO 361
 |||||||

RESULT 3
 ID Q9YK34 PRELIMINARY; PRT: 823 AA.
 AC Q9YK34;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE SIR2. PROTEIN.
 GN SIR2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodet A., Gong F., Gotteff J.H., Gu Z., Guan P., Harris K.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jaisai M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusker D.R., Paclob J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Stiden-Klamos I., Simpson M., Skupski M.P., Slat T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003639; AAP53248.1; -
 DR FLBASE, FBgn0024291; Sir2.
 DR INTERPRO: IPR003000; -
 DR PFM: PFM02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 91836 MW; 3DD1EBA975A263D3 CRC64;

Query Match 84.7%; Score 471; DB 5; Length 823;
 Best Local Similarity 80.2%; Pred. No. 1.8e-43;
 Matches 85; Conservative 14; Mismatches 7; Indels 0; Gaps 0;
 Oy 1 IIVLTGAGVSVCIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 60
 |||||||
 Db 224 IIVLTGAGVSVCIPDRSRDGIYARLAVDPDLPDQAMFDIEYFKDRPFFKFAKE 283
 |||||||
 Oy 61 IYPGQFQPSLCHKFTALSDKKGKLLRNTONTIDLEQVAGIORTIIO 106
 |||||||
 Db 284 IYPGQFQPSLCHKFTALSDKKGKLLRNTONTIDLEQVAGIORTIIO 329
 |||||||

RESULT 4
 ID Q96505 PRELIMINARY; PRT: 823 AA.
 AC Q96505;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE SIR2.
 GN SIR2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Astrom S.U., Rine J.;
 RL Submitted (May-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AF068758; AAC79684.1; -
 DR INTERPRO: IPR003000; -
 DR PFM: PFM02146; SIR2; 1.
 SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A46235E7F CRC64;
 Query Match 83.3%; Score 463; DB 5; Length 823;
 Best Local Similarity 79.6%; Pred. No. 1.4e-42;

Matches 86; Conservative 13; Mismatches 7; Indels 2; Gaps 1;

QY 1 IIVLTGAGVSCGIPFRSDGIYARLAVDPDPQAMFDEYFRKDPPEFFAK 60
 Db 222 IIVLTGAGVSCGIPFRSTNGIYARLADFPDPQAMFDPYKRPPEFFAK 281

QY 61 IYGG-0FOPSLCHKFTALSDKEGKLRYNTONIDTLEOVAGIORIQ 106
 Db 282 IYGGFOPSPCHRFIKMETKCKLLRYNTONIDTLEOVAGIORVIE 329

RESULT 5
 ID 021921 PRELIMINARY; PRT; 607 AA.
 AC 021921;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE R1A8.4 PROTEIN.
 GN R1A8.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chordata; Rhabditiida; Rhabditidae;
 OC Rhabditiidae; Pelodertinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cummings P.,
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; Pubmed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten T., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL:Z70310; CA94364.1; -;
 DR INTERPRO: IPR003000; -;
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 607 AA; 68765 MW; 71C720BABA4DC8BD CRC64;

Query Match 73.7%; Score 410; DB 5; Length 607;
 Best Local Similarity 69.8%; Pred. No. 6, 6e-37;
 Matches 74; Conservative 16; Mismatches 16; Indels 0; Gaps 0;

QY 1 IIVLTGAGVSCGIPFRSDGIYARLAVDPDPQAMFDEYFRKDPPEFFAK 60
 Db 148 IIVLTGAGVSCGIPFRSKDGIYARLSEFPDPPTAMFDRYFRNPAPYFAR 207

QY 61 IYGGFOPSPCHRFIKMETKCKLLRYNTONIDTLEOVAGIORIQ 106
 Db 208 IYGGFOPSPCHRFIKMETKCKLLRYNTONIDTLEOVAGIORVIE 253

RESULT 6
 ID 095889 PRELIMINARY; PRT; 254 AA.
 AC 095889;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE SMILAR TO RAT 5E5 ANTIGEN (FRAGMENT).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96207227; Pubmed=8619474;
 RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.,
 RT "A 'double adaptor' method for improved shotgun library
 RT construction.";
 RL Anal. Biochem. 236:107-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97264341; Pubmed=9110174;
 RA Yu W., Andersson B., Morley K.C., Muzny D.M., Ding Y., Liu W.,
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.,
 RL "Large-scale concatenation cDNA sequencing.";
 RN Genome Res. 7:353-358(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mel G., Yu W., Gibbs R.A.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF131800; AAD20046.1; -;
 DR INTERPRO: IPR003000; -;
 DR PFAM: PF02146; SIR2; 1.
 FT NON TER 1
 SQ SEQUENCE 254 AA; 28577 MW; 5AFAE01758E476E3 CRC64;

Query Match 54.2%; Score 301.5; DB 4; Length 254;
 Best Local Similarity 55.8%; Pred. No. 2, 1e-25;
 Matches 58; Conservative 19; Mismatches 24; Indels 3; Gaps 2;

QY 1 IIVLTGAGVSCGIPFRS-RDGIYARLAVDPDPQAMFDEYFRKDPPEFFAK 59
 Db 62 VICVAGAGISASGIPFRSPSTGLYDNL--EKYHLPPEAIFSIYFKKHPEFFALAK 119

QY 60 EYGGFOPSPCHRFIKMETKCKLLRYNTONIDTLEOVAGIOR 103
 Db 120 EYGGFOPSPCHRFIKMETKCKLLRYNTONIDTLEOVAGIOR 163

RESULT 7
 ID 09UNTO PRELIMINARY; PRT; 352 AA.
 AC 09UNTO;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE SILENCING INFORMATION REGULATOR 2-LIKE PROTEIN.
 GN STR2L.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99321810; Pubmed=10393250;
 RA Atshar G., Murnane J.P.,
 RT "Characterization of a human gene with sequence homology to
 RT Saccharomyces cerevisiae SIR2.";
 RL Gene 234:161-168(1999).
 DR EMBL: AF095714; AAD45971.1; -;
 DR INTERPRO: IPR003000; -;
 DR PFAM: PF02146; SIR2; 1.
 SQ SEQUENCE 352 AA; 39515 MW; FFED07DEF9E3416A CRC64;

Query Match 54.2%; Score 301.5; DB 4; Length 352;
 Best Local Similarity 55.8%; Pred. No. 3, 1e-25;
 Matches 58; Conservative 19; Mismatches 24; Indels 3; Gaps 2;

QY 1 IIVLTGAGVSCGIPFRS-RDGIYARLAVDPDPQAMFDEYFRKDPPEFFAK 59
 Db 42 VICVAGAGISAGIPFRSPSTGLYDNL--EKYHLPPEAIFSIYFKKHPEFFALAK 99

0y	60	EEYGFQFSPKSHKFIALSDKSGKILRYNTQINIDLEAVQI	103
		::: :: ::: :: ::	
Db	100	ELYPQGFPTTICHYFMRLKKDGLLRCYTONIDTLRIAGLEQ	143
RESULT	8		
09Y6E9			
ID	09Y6E9	PRELIMINARY;	PRN; 373 AA.
AC	09Y6E9;		
DT	01-NOV-1999	(TREMBLrel. 12, Created)	
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	SIRTUIN TYPE 2.		
GN	SIRT2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TESTIS;		
RC	MEDLINE=99310604; Pubmed=10381378;		
RT	Frye R.A.;		
RT	"Characterization of five human cDNAs with homology to the yeast SIRT2		
RT	gene: Sir2-like proteins (sirtulins) metabolize NAD and may have		
RT	protein ADP-ribosyltransferase activity.";		
RL	Biochem. Biophys. Res. Commun. 260:273-279(1999).		
DR	EML; AF083107; AAD0850.1; -.		
DR	INTERPRO: IPR003000; -.		
DO	SEQUENCE 373 AA; 41497 MW; C0985F3D81D5D815 CRC64;		

```

Query Match          54.2% Score 301.5; DB 4; Length 373;
Best Local Similarity 55.8%; Pred. No. 3.3e-25;
Matches 58; Conservative 19; Mismatches 24; Indels 3; Gaps

OY      1 IIVLTGAGVSVCIPDRS-RDSIYARLADFFDLDPDQAMEDIYFRKDPPEFFAK 59
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      63 VICLGAGISHSAGIPDRSPSTGLNDL--EKHYLHPREAIPISTFKKHPEFFALAK 120
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      60 EIVEGFQPSLCNKFIALSDKEGLLNRYTQNIDTLSEQVGIQR 103
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121 ELVPGQFPTICHYFMRLDKGLLRKYQTNIIDLRIAGLEQ 164
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT   9
O94066    PRELIMINARY; PRT; 331 AA.
AC O94066;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE TRANSCRIPTION REGULATORY PROTEIN.
GN CA9CA.08C.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC anaerophilic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Murphy L., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Barrett B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=97435544; PubMed=9290243;
```

RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT "A Candida albicans genome project: cosmid contigs, physical mapping
RT and gene isolation";
RL Fungal Genet. Biol. 21:308-314 (1997).
DR EMBL; AL033503; CA22018.1; "-".
DR INTERPRO; IPR003000; "-".
DR PFAM; PF02146; SIR2; 1.
SQ SEQUENCE 331 AA; 37509 MM; 8748DBB4BED79D10 CRC64;

Query Match	49.6%	Score 275.5	DB 3	Length 331
Best Local Similarity	51.5%	Pred. No. 2.1e-22		
Matches 53, Conservative	20	Mismatches 27	Indels 3	Gaps 2

[illegible]

RESULT	10
09Y6E8	
ID	09Y6E8
AC	09Y6E8
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	SIRTUIN TYPE 3.
OS	SIRT3.
CS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=TESTIS;
RX	MEDLINE=99310604; PubMed=10381378;
RA	Frye R.A.;
RT	"Characterization of five human cDNAs with homology to the yeast SIRT
RT	gene: Sirt2-like proteins (sirtulins) metabolize NAD and may have
RT	protein ADP-ribosyltransferase activity.";
RL	Biochem. Biophys. Res. Commun. 260:273-279(1999).
DR	EMBL: AF083108; AAD40851.1; -
DR	INTERPRO: IPR003000; -
SO	SEQUENCE 399 AA; 43573 MW; 4BABBD3AC5FC7901 CRC64;

Query Match	47.4%;	Score 263.5;	DB 4;	Length 399;
Best Local Similarity	50.0%;	Pred. No. 5.5e-21;		
Matches	51;	Conservative	22;	Mismatches 26; Indels 3; Gaps 2;

QY	1	IIIVLTGAQVSVSCGIPDRRS-RDGIYALANDPFLPDPQAMFIDEYRKDRPFFKFAK	59
Db	140	VVVVGAGDISPSPGIPDRSPSSGILSYNL--QOYDLPPEALFELPFFHNPKEPFTLAK	197
QY	60	EIYPGQFQPSLCHKRIALSDKEGKLLRYNTONIDPLEGVAGI	101
Db	198	EIYPGNKYKPNVTHYFLRLHDKGLLRLYQTQINDIGLEKRSGI	239

RESULT	11
Q9USN7	
ID	Q9USN7
AC	Q9USN7
	PRELIMINARY;
	PRT;
	332 AA.
DT	01-MAY-2000 (Tremblere1.13, Created)
DT	01-MAY-2000 (Tremblere1.13, Last sequence update)
DT	01-OCT-2000 (Tremblere1.15, last annotation update)
DE	POTATIVE SIR2-LIKE TRANSCRIPTIONAL REGULATORY PROTEIN.
	SPEC132.02.
CS	Schizosaccharomyces pombe (Fission yeast).

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 10:51:38 ; Search time 137.23 Seconds
(without alignments)
629.470 Million cell updates/sec

Title: US-09-461-580A-26
Perfect score: 3854
Sequence: 1 MADEVALALQAGSPSA...NEALATROELTDVNPSPDKS 737

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORNITHINE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3854	100.0	737	11	Q9QXG8	Q9QXG8 mus musculu
2	3313	86.0	747	4	Q9YGE0	Q9YGE0 homo sapien
3	1169	30.3	823	5	Q9YK34	Q9YK34 drosophila
4	1157.5	30.0	823	5	Q9YK34	Q9YK34 drosophila
5	923.5	24.0	607	5	Q21921	Q21921 caenorhabdi
6	605.5	15.7	471	3	Q94640	Q94640 schizosacch
7	556.5	14.4	373	4	Q9YGE9	Q9YGE9 homo sapien
8	555.5	14.4	352	4	Q9UNT0	Q9UNT0 homo sapien
9	538	14.0	331	3	Q94066	Q94066 candida alb
10	513.5	13.3	332	3	Q9USN7	Q9USN7 schizosacch
11	508	13.2	399	4	Q9YGE8	Q9YGE8 homo sapien
12	493	12.8	254	4	Q9YGE8	Q9YGE8 homo sapien
13	406	10.5	351	5	Q96670	Q96670 tripanosoma
14	405.5	10.5	373	4	Q9PGC8	Q9PGC8 homo sapien
15	349	9.1	247	2	Q07595	Q07595 bacillus su
16	346	9.0	246	2	Q9WY00	Q9WY00 thermotoga
17	330.5	8.6	253	1	Q30124	Q30124 archaeoglob
18	295	7.7	245	1	Q28597	Q28597 archaeoglob
19	285.5	7.4	247	1	Q9YB13	Q9YB13 aeropyrum p

20	282.5	7.3	249	1	Q58669	Q58669 pyrococcus
21	267.5	6.9	250	1	Q9UZE7	Q9UZE7 pyrococcus
22	264	6.9	243	2	Q53700	Q53700 staphylococ
23	262	6.8	310	4	Q9YGE6	Q9YGE6 homo sapien
24	256	6.6	415	3	Q9UR39	Q9UR39 schizosacch
25	251.5	6.5	299	4	Q9NXX8	Q9NXX8 homo sapien
26	245	6.4	239	2	Q67919	Q67919 aquifex aeo
27	237	6.1	314	4	Q9YGE7	Q9YGE7 homo sapien
28	233	6.0	326	4	Q43346	Q43346 homo sapien
29	230	6.0	237	2	Q9KEE5	Q9KEE5 bacillus ha
30	220	5.7	287	5	Q20480	Q20480 caenorhabdi
31	216.5	5.6	299	2	Q9RL35	Q9RL35 streptomyce
32	216	5.6	377	4	Q9UOD1	Q9UOD1 homo sapien
33	212	5.5	355	4	Q9NRC7	Q9NRC7 homo sapien
34	211	5.5	237	2	Q06549	Q06549 mycobacteri
35	210	5.4	259	2	Q9KRX4	Q9KRX4 vibrio chol
36	210	5.4	279	2	P75960	P75960 escherichia
37	199.5	5.2	208	2	Q9ZAB8	Q9ZAB8 actinobacti
38	197.5	5.1	234	2	Q9ZJW8	Q9ZJW8 helicobacte
39	190	4.9	287	5	Q20481	Q20481 caenorhabdi
40	187.5	4.9	233	2	Q9PNP1	Q9PNP1 campylobact
41	186.5	4.8	483	10	Q9XGT1	Q9XGT1 oryza sativ
42	186	4.8	227	2	Q9JN05	Q9JN05 campylobact
43	186	4.8	3534	12	Q39266	Q39266 equine herp
44	185.5	4.8	246	2	Q9RYD4	Q9RYD4 deinococcus
45	185	4.8	320	5	Q9U1A6	Q9U1A6 leishmania

ALIGNMENTS

RESULT	ID	Q9QXG8	PRELIMINARY:	PRT:	737 AA.
Q9QXG8	Q9QXG8	Q9QXG8			
AC	Q9QXG8	Q9QXG8			
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, last annotation update)			
DE	STR2ALPHA PROTEIN.				
GN	SIR2ALPHA.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SYRAIN-SWISS WEBSTER/NIH;				
RA	Imai S., Armstrong C.M., Guarente L.;				
RT	"Silencing and ageing protein Sir2 is an NAD-dependent histone				
RT	deacetylase."				
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AF214646; AAF24983.1; -				
DR	HSSP: P04002; IATF.				
DR	INTERPRO: IPR003000; -				
PR	PRAM: PR02146; SIR2; 1				
DR	SEQUENCE 737 AA; 80371 MW; 7F15625E29433119 CRC64;				
SO	SEQUENCE				
Query Match	100.0%; Score 3854; DB 11; Length 737;				
Best Local Similarity	100.0%; Pred. No. 4, 1e-263;				
Matches 737; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MADEVALLAQAGSPSA...NEALATROELTDVNPSPDKS 737			
DB	1	MADEVALLAQAGSPSA...NEALATROELTDVNPSPDKS 737			
QY	61	CEAASAAAPALMREAA...GAASAREAPATVAVGDSGSGLRREPAADFDDEEE 120			
DB	61	CEAASAAAPALMREAA...GAASAREAPATVAVGDSGSGLRREPAADFDDEEE 120			
QY	121	DEAASAAAPALMREAA...GAASAREAPATVAVGDSGSGLRREPAADFDDEEE 180			
DB	121	DEAASAAAPALMREAA...GAASAREAPATVAVGDSGSGLRREPAADFDDEEE 180			

```

QY 181 00HLMIGTDPRTILKDLPEITIPPELDMTLMQIVINILSEPPKRRKKRDINTIEDAVK 240
D 181 00HLMIGTDPRTILKDLPEITIPPELDMTLMQIVINILSEPPKRRKKRDINTIEDAVK 240
QY 241 LLOECKKIIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDQAMFIEYFKKDRP 300
D 241 LLOECKKIIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPDQAMFIEYFKKDRP 300
QY 301 FFFKFAKEIYPGOPFSLCHKFIKALSDKEGKILRNNTQNTIDLEQVAGIQRILQCHGSEAT 360
D 301 FFFKFAKEIYPGOPFSLCHKFIKALSDKEGKILRNNTQNTIDLEQVAGIQRILQCHGSEAT 360
QY 361 ASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYFGENLPEQFHRAMKY 420
D 361 ASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYFGENLPEQFHRAMKY 420
QY 421 DKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQILINREPLPHLFDVLLGDCVITNE 480
D 421 DKDEVDLLIVIGSSLKVRPALIPSSIPHEVPQILINREPLPHLFDVLLGDCVITNE 480
QY 481 LCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLHISDSSSEPTVPD 540
D 481 LCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLHISDSSSEPTVPD 540
QY 541 SSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVENINENPDFKAVGSSADK 600
D 541 SSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVENINENPDFKAVGSSADK 600
QY 601 NERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIFHGAEVYSDSEDDVLSSSS 660
D 601 NERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIFHGAEVYSDSEDDVLSSSS 660
QY 661 CGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECAAGSGGFADGDDQVNEA 720
D 661 CGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECAAGSGGFADGDDQVNEA 720
QY 721 IATROELTDVNPYSDKS 737
D 721 IATROELTDVNPYSDKS 737

```

```

Query Match 86.0%: Score 3313; DB 4: Length 747;
Best Local Similarity 85.4%: Pred. No. 5e-225;
Matches 645; Conservative 35; Mismatches 49; Indels 26; Gaps 10;

QY 1 MADEVALALQAAQSPSAAA--MEASQPADPELRKRRPRDGGGLRSFGERSAA-----V 54
D 1 MADEVALALQPGGSPSAGADREMAASPAGEPLRKRPRDGGGLRSPGEGGAAPEREV 60
QY 55 AAPAAGCEAASAAAPALMR--EAAGAAASAAREAPATAVAGDGGNGSLR---REPPAA 109
D 55 AAPAAGCEGAAA--AALMREAEMAAAGGEQAATAAGGEGNGGLOGPSREPLA 117
QY 110 DDF-----DDEGEEDDEAAAAAAGYRONLLTDLGTLNGFSCSDDDDRSHASS 165
D 110 DNLVEDDDDEGEDEE-----AAAAAIGYRONLLFGDEITINGFSCSDDEBRASHASS 173
QY 166 SDWTPRPRIQPYTFVQOHLMITGDPRTILKDLPEITIPPELDMTLMQIVINILSEPPK 225
D 166 SDWTPRPRIQPYTFVQOHLMITGDPRTILKDLPEITIPPELDMTLMQIVINILSEPPK 225
QY 174 SDWTPRPRIQPYTFVQOHLMITGDPRTILKDLPEITIPPELDMTLMQIVINILSEPPK 233
D 174 SDWTPRPRIQPYTFVQOHLMITGDPRTILKDLPEITIPPELDMTLMQIVINILSEPPK 233
QY 226 RKKRDIWITIEDAVKALDCECKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPD 285
D 226 RKKRDIWITIEDAVKALDCECKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPD 285
QY 234 RKKRDIWITIEDAVKALDCECKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPD 293
D 234 RKKRDIWITIEDAVKALDCECKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDPDLPD 293
QY 286 QAMFDIEYFRKDPREFKFAKEIYPGOPFSLCHKFIKALSDKEGKILRNNTQNTIDLEQV 345
D 286 QAMFDIEYFRKDPREFKFAKEIYPGOPFSLCHKFIKALSDKEGKILRNNTQNTIDLEQV 345
QY 346 AGIQRILIOCHGSFATASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYF 405
D 346 AGIQRILIOCHGSFATASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYF 405
QY 354 AGIQRILIOCHGSFATASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYF 413
D 354 AGIQRILIOCHGSFATASCLICKYKVDCEAVRGDIFNQVYPRCPADPEPLAIKMKPEIYF 413
QY 406 FGENLPEQFHRAMKDKQEVLLIYIGSSLKVRPALIPSSIPHEVPQILINREPLPHL 465
D 406 FGENLPEQFHRAMKDKQEVLLIYIGSSLKVRPALIPSSIPHEVPQILINREPLPHL 465
QY 414 FGENLPEQFHRAMKDKQEVLLIYIGSSLKVRPALIPSSIPHEVPQILINREPLPHL 473
D 414 FGENLPEQFHRAMKDKQEVLLIYIGSSLKVRPALIPSSIPHEVPQILINREPLPHL 473
QY 466 FDVELLGGCDVITINELCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLH 525
D 466 FDVELLGGCDVITINELCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLH 525
QY 474 FDVELLGGCDVITINELCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLH 533
D 474 FDVELLGGCDVITINELCHRLGGEYAKLCNPVKLSEITEKPPRPOKELVHSELPPPLH 533
QY 526 ISEDSSEPTVPDSSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVENI- 583
D 526 ISEDSSEPTVPDSSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVENI- 583
QY 534 VSEDSSEPTVPDSSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVESIA 592
D 534 VSEDSSEPTVPDSSVIATLVDOATNNNVNNDLEVESSESCVEKEKPEVQTSRNVESIA 592
QY 584 -NVEPDPKAVGSSADNERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIF 642
D 584 -NVEPDPKAVGSSADNERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIF 642
QY 593 EOMENPDLKNVSSYSGEKNERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIF 652
D 593 EOMENPDLKNVSSYSGEKNERTSVAETVRKCMFNRLAKEQISKRLEGNOYLFVPPNRYIF 652
QY 643 HGAEYVSDSEDDVLSSSCGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECA 702
D 643 HGAEYVSDSEDDVLSSSCGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECA 702
QY 653 HGAEYVSDSEDDVLSSSCGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECA 712
D 653 HGAEYVSDSEDDVLSSSCGNSDSTGQSPSLSEPLEDESETEEFYNGLEDTERPECA 712
QY 703 GGSFGADGDDQEVVNEAIATROELTDVNPYSDKS 737
D 703 GGSFGADGDDQEVVNEAIATROELTDVNPYSDKS 737
QY 713 GGAGGTGDDQDEAINEAISVKQEVTDNMNPSNKS 747
D 713 GGAGGTGDDQDEAINEAISVKQEVTDNMNPSNKS 747

```

```

RESULT 2
Q9Y6F0 PRELIMINARY; PRT; 747 AA.
AC Q9Y6F0:
DT 01-NOV-1999 (Tremblrel. 12. Created)
DT 01-OCT-2000 (Tremblrel. 15. Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15. Last annotation update)
DE SIRTUIN TYPE 1.
GN SIRTUIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RC SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RC MEDLINE=99310604; PubMed=10381378;
RA Frye R.A.;
RT "Characterization of Five Human cDNAs with Homology to the yeast SIRT2
RT Gene; SIRT-Like Proteins (Sirtuins) Metabolize NAD and May Have
RT Protein ADP-Ribosyltransferase Activity."
RL Biochem. Biophys. Res. Commun. 260:273-279(1999).
DR EMBL: AF083106; A040849.2;
DR INTERPRO: IPR003000;
DR PFM; PFO2146; SIRT2; 1.
SQ SEQUENCE 747 AA; 81680 MW; 2D3BEA6D73DA229F CRC64;

```

```

RESULT 3
Q9VK34 PRELIMINARY; PRT; 823 AA.
AC Q9VK34:
DT 01-MAY-2000 (Tremblrel. 13. Created)
DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15. Last annotation update)
DE SIRT2 PROTEIN.
GN SIRT2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN-BERKELEY:
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman I.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003639; AAF53248.1;
 DR FLYBASE: FBgn0024291; S1R2.
 DR INTERPRO: IPR003000;
 DR PFM: PF02146; S1R2; 1.
 SQ SEQUENCE 823 AA; 91836 MW; 3DD1EBA975A263D3 CRC64;

Query Match 30.3%; Score 1169; DB 5; Length 823;
 Best Local Similarity 38.8%; Pred. No. 4, 9e-74;
 Matches 293; Conservative 100; Mismatches 206; Indels 156; Gaps 25;
 DB 57 AAGCEAASAAAPALMREA-----GAASAREAPATAVAGGDSGLRREPR-----107
 DB 42 ASTSTAEAEATATTPATSELAKANGELKTLAREOEIGANLEHKTPTKS 101
 QY 108 AADDFDDDEGEDEEAAAAAAAAAIGYRDNLITDGLTNGFHSCESDDDRTSHASS- 166
 DB 102 MCEDEDEDEEEDDEE-----EDD-----EKGITGTSENEDDESSNCSSSV 145
 QY 167 --DWTFRPRIGTYTVOQLMIGTDPRTILKDLPE--TIPPELDDMTLMQIVINISL 222
 DB 146 EDPWKLR-----WLQREFYTGVRVROYIASIMPHFATGLADTDSVLMWYLAHLN 198
 QY 223 PRKRKRKDKINTEDTAVKLLQECKKITIVTGAGVSVCIGPDRSRDGIYARLAVFPDL 282
 DB 199 PRKRKLASVNTFDVYISLVKKSQKITIVLTGAGVSVCIGPDRSRNGIYARLAHPDL 258
 QY 283 PDQAMFDEIEFRKDRPFPEKFEKLEYPOQFOPSLCHKEFTALSDKCKLLRNTQNDTL 342
 DB 259 PDQAMFDEIEFRKDRPFPEKFEKLEYPOQFOPSLCHKEFTALSDKCKLLRNTQNDTL 318

QY 343 EQVAGIQRILQCHGSEFATASCLICKYKYVDCAEVRGDIQNVVPRCRPC-----391
 DB 319 ENAGIQRVIECHGSPSTASCTKCRKCNADLRADIFQRIIPVCGQCPNNQGSVDASV 378
 QY 392 --ADEPL-----AIKPEIIVEFGENLPQGFHRAKMYKDEVDLLIVIGSSLKVRVALI 443
 DB 379 AVTEELRLQVLENGIMKPDIVFEFEGELPDEHYATVMATDKVCDLLIVIGSSLKVRVAHI 438
 QY 444 PSSIPHEVQIILNRPDLPHLPDVELLDCCYIINELCHRLGGE---YAKLCNPKVS 500
 DB 439 PSSIPHEVQIILNRPDLPHLPDVELLDCCYIINELCHRLGGE---YAKLCNPKVS 498
 QY 501 ETEKPRROKELVHSELPPPLHISEDSSPER-----TVPDSSYATLVDA 551
 DB 499 ESKELMP-FEHSNHHH--HLLHRRHSSSESRQSOLDTQSIKSNSSADYILGSG 554
 QY 552 TANNVDELVSSESC-----VE-----EKKP-EVOT 576
 DB 555 TCSG-SGFESSFSCGKRSRTAAEAARIKTDILVELNETTALSCDRGLGEGQPTVES 613
 QY 577 SRV-----ENINVENPDRKAVGSSYADN---ERTSVAEYKCKPNRLAKEQISRL 628
 DB 614 YRLSIDSSKDSGLEQCDNEAPPSYVRPSNLVQETVAPSLPIPOGKQRTAERLQ 673
 QY 629 -GNQYLFVPNRYIFHGAEV-----YSDSEDDVLSS-----SCGSN---SDSGTC 670
 DB 674 PGTFYSHNTNYSIVFPGAQVFMNDYSDDDDEEESHNRHSDLFQNGVHNKDDDEDAC 733
 QY 671 Q-----SPSLE-----EPLEDES 683
 DB 734 DINAVALSPPLPSLEAHITVDIVNGSNPLPSS 768
 RESULT 4
 ID 096505 PRELIMINARY; PRT; 823 AA.
 AC 096505
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE S1R2.
 GN S1R2.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Astrom S.U., Rine J.;
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF068758; AAC79684.1;
 DR INTERPRO: IPR003000;
 DR PFM: PF02146; S1R2; 1.
 SQ SEQUENCE 823 AA; 92023 MW; 5EF09F1A46235E7F CRC64;

Query Match 30.0%; Score 1157.5; DB 5; Length 823;
 Best Local Similarity 39.3%; Pred. No. 3, 1e-73;
 Matches 289; Conservative 102; Mismatches 210; Indels 135; Gaps 25;
 DB 52 AAVAPAAACEAASAAAPALMREAAGAAASAREAPATAVAGGDSGLRREPR-----107
 DB 42 ASTSTAEAEATATTPATSELAKANGELKTLAREOEIGANLEHKTPTKS 98
 QY 108 AADDFDDDEGEDEEAAAAAAAAAIGYRDNLITDGLTNGFHSCESDDDRTSHASS 166
 DB 99 SMGEDEDEEEDDEE-----EDD-----EKGITGTSENEDDESSNCSSSV 142
 QY 167 --DWTFRPRIGTYTVOQLMIGTDPRTILKDLPE--TIPPELDDMTLMQIVINISL 221
 DB 143 VEDPWKLR-----WLQREFYTGVRVROYIASIMPHFATGLADTDSVLMWYLAHLN 195


```

Db 130 QUIDLERLAGLEODLVFAHGTFTSHCVSASCHREYPLSMWKKETSEVTPKEDCO- 188
Oy 393 DEPLAIKPEIYFGEFPEOFHRAKMYDKDEVDLLIVIGSSLAKRPVLISSIPHEVP 452
Db 169 ----SLVRDIFYFFGESLPARFSCMQSDFLKVDLLVWGTSLQVPPFSLISKAPLSTP 244
Oy 453 QILINREPL-----PHL-----HFDVLLGDCDVIINELCHRLG----- 486
Db 245 RLLINKKAGQSDPFLGIMIGLGGMDPSKKAYRDVAMLGCDGCLALABELGKKEL 304
Oy 467 -----GEYAKLCC-----NPVKLSEITEKPPROKELVHLSELPPPLIHSESSSPE 534
Db 305 EDLVRREHNASIDAGAGVNP-STASAPKSPPPAK-----DEARTTE 347
Oy 535 RTVPQ 539
Db 348 REKPO 352

```

RESULT 9

```

O94066 PRELIMINARY; PRT; 331 AA.
ID O94066
AC O94066
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE TRANSCRIPTION REGULATORY PROTEIN.
GN CA49C4.08C.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC anamorphic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Murphy L., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA MEDLINE=97435544; PubMed=9290243;
RT Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT "A Candida albicans genome project: cosmid contigs, physical mapping,
RT and gene isolation."
RL Fungal Genet. Biol. 21:308-314 (1997).
DR EMBL; AL033503; CAA22018.1; -.
DR INTERPRO: IPR003000; -.
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 331 AA; 37509 MW; 8748DFB4ED79D10 CRC64;

```

Query Match 14.0%; Score 538; DB 3; Length 331;
 Best Local Similarity 34.8%; Pred. No. 3.7e-30;
 Matches 128; Conservative 69; Mismatches 117; Indels 54; Gaps 11;

```

Oy 205 PELDMLQIVINILSEPPKRRKRDINTEDAVKLLQECKRIIVLGAGVSCGIPD 264
Db 2 PSLDI-----LKPAEAVK-----NGKKVFFPNAGISTGAGIPD 37
Oy 265 FRSRD-GIYARLAVFPDLPQAMFDIEYFRKDRPFPEFKKELYPGQFOPSLCHKFLA 323
Db 38 FRSPGTGYANLAK--LNLPEAAVAFDIDFEKEDPKPFYTLAEELYPGNFAPTKHFHFK 95
Oy 324 LSDKGKLLRNTYONIDTLERAGVAGIO--RILOCHGSFATASCLICKYKDCAVRGDIEN 381
Db 96 LLODGSILKRYTTONIDTLERLAGVEDKTYIEAHGSFASNNCHVDCHKEKTELTLYMKD 155

```

```

Oy 382 QVPRCPRCPADDEPLAIKPEIYFGENLPEQFHRAKMYDKDEVDLLIVIGSSLKVRPA 441
Db 156 KIPSCQHC-----GYVPDIFYFGEGLPVKFFDLMEDDCEDVEAIVAGTSLVFPFA 210
Oy 442 LIPSPHEVPQILINREPLPHLF-----DVLLGDCDVIINELCHRLGSEYAKLCCNP 496
Db 211 SLPEVNNKCLVNLVKNKEVGVTFKHEPRKSDIIALHDCDVIKELCTLLGLD----- 262
Oy 497 VKLSEITEKPPROKELVHLSELPPPLIHSESSSPERTVPQDSSVIATLVD--QATNN 555
Db 263 DKLNEYEK-----EKIKSKAETKETKMEIEDKLKEAHLKEDKH--TTKYDNKEKOND 316
Oy 556 VNDLEVSE 563
Db 317 ANDKELEQ 324

```

RESULT 10

```

O9USN7 PRELIMINARY; PRT; 332 AA.
ID O9USN7
AC O9USN7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE SIR2-LIKE TRANSCRIPTIONAL REGULATORY PROTEIN.
GN SPC132.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetales; Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Aert R., Volckaert G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121807; CAB58129.1; -.
DR INTERPRO: IPR003000; -.
DR PFAM: PF02146; SIR2; 1.
SQ SEQUENCE 332 AA; 37915 MW; 936A08D240C8BCFF CRC64;

```

Query Match 13.3%; Score 513.5; DB 3; Length 332;
 Best Local Similarity 35.8%; Pred. No. 2e-28;
 Matches 126; Conservative 63; Mismatches 112; Indels 51; Gaps 10;

```

Oy 225 IEDAVKLLQE--CKKIIVLTGAGVSCGIPERSRD-GIYARLAVDPDLPQAMFDI 291
Db 15 LEKVASLIEKGKVKRICVAVGAGISTAGIDPERSPETGIVNNL--QRENLYAEAVEDL 72
Oy 292 EYFRKDRPFPEFKAKIYPGQFOPSLCHKFLALSDKEGKLLRNTYONIDTLERAGVAGI--Q 349
Db 73 SYFRKNRPREFYLAHMEKRYTYHYFIRLLDKRLQCYONIDTLERLAGVDPK 132
Oy 350 RILQCHGSFATASCLICKYKDCAVRGDIENQVPRCPRCPADDEPLAIKPEIYFGEN 409
Db 133 ALIEAHGSFOYSRCIECEYEMETEVFRACIMQKQVPCNSCK-----GLIKPMIYFVGE 187
Oy 410 LPEQFHRAKMYDKDEVDLLIVIGSSLKVRPVALLIPSIPHEVPQILINREP--LPHLF 466
Db 168 LPMREFFHEKTKYCDMALVIGTSLVHPFADLEIYVNNKQKQVRLINERPGDFGERKK 247
Oy 467 DVLLGDCDVIINELCHRLG--GEYAKLCCNPVKLSEITEKPPROKELVHLSELPPPL 524
Db 248 DMIILGDCDSQVRAKCLKLLGWSDELEKLLIDISVE--TLTE-----ISL----- 290
Oy 525 HISEDSSPERTVPQDSSVIATLVDQATNNVNDLEVSSESCVEEKPQEVOT 576
Db 291 -----SVDSITIEKNASEQKKDKNVSNPFTIEEKKKDEVT 325

```

RESULT 11
 O9Y6E8 PRELIMINARY; PRT; 399 AA.

AC 09Y6E8;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, last annotation update)
 DE SIRTUIN TYPE 3.
 GN SIRT3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=99310604; PubMed=10381378;
 RA Fyfe R.A.;
 RT "Characterization of five human cDNAs with homology to the yeast SIRT3 gene: Sirt2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity."
 RT Blochem. Biophys. Res. Commun. 260:273-279(1999).
 RL EMBL: AF083108; AAD40851.1; -
 DR INTERPRO: IPR003000; -
 DR SEQUENCE 399 AA; 43573 MW; 4B8BD3AC5FC7901 CRC64;
 SQ
 Query Match 13.2%; Score 508; DB 4; Length 399;
 Best Local Similarity 40.4%; Pred. No. 6, 2e-28;
 Matches 107; Conservative 60; Mismatches 76; Indels 22; Gaps 8;
 QY 234 TIEDAVKL--QECKIIVLTGAGVSVSGIPDRS--RDGIYARLAVDPDLPDPOAMD 290
 DB 124 SLOVAVELIRACQVRVVMVAGISIPDRSPSGISYML--QYVLLPPEALFE 181
 QY 291 IEYRKRPPPEFKAKETIQEQFQPSLCHKFIASDKEGKLIRNTQNTIDLEOVAGI-- 348
 DB 182 LPFEFHNPKPEFLAKELYPENYKRNWYHFLRLDHLKGLRLTYONIDLERVSGIPA 241
 QY 349 QRILOCHGSPFATSCILCKYKVDCEAVRGDIENGVPRCPDPADEPLAIKPEIVFGE 408
 DB 242 SKLVEANGTFRSATCTVCQRFPGEDIRADVADRVRPCVC-----TGVKRPDIVFGE 296
 QY 409 NLPRQF--HRAKMYKDEVDLLIIGSSLVKRYVALIPSSIPHEVPOILLINREPLPLHF 466
 DB 297 PLPQRFILH--VYDFPADLLILIGTSLVEYEPASLAEAVRSSVPRLLINRDLVGLAW 353
 QY 467 -----DVELLGDCCVIITNELCHRI 486
 DB 354 HPRSRDVAQLGDVYGVESIVELLG 378
 RESULT 12
 ID 095889 PRELIMINARY; PRT; 254 AA.
 AC 095889;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
 DE SMTILAR TO RAT S55 ANTIGEN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96207227; PubMed=8619474;
 RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
 RT "A 'double adaptor' method for improved shotgun library construction."
 RT Anal. Biochem. 236:107-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9726341; PubMed=9110174;
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W., Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;

RT "Large-scale concatenation cDNA sequencing."
 RL Genome Res. 7:353-358(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mei G., Yu W., Gibbs R.A.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF131800; AAD20046.1; -
 DR INTERPRO: IPR003000; -
 DR PRAM: PF02146; SIRT2; 1.
 DR NON_TER 1
 SQ SEQUENCE 254 AA; 28577 MW; 5AEAE01758E476E3 CRC64;
 Query Match 12.8%; Score 493; DB 4; Length 254;
 Best Local Similarity 42.4%; Pred. No. 3, 8e-27;
 Matches 103; Conservative 56; Mismatches 70; Indels 14; Gaps 6;
 QY 206 ELDDMTLMQIYINILSEPPKRRKRKRDINTIEDAVKLLD--ECKRIIVLTGAGVSVSGIP 263
 DB 18 EADMDLNLFLSQTLSGSKERLDELTLGVARVMQSEGRVICYAGISITSAGIP 77
 QY 264 DFRS--RDGIYARLAVDPDLPDPOAMDIEYFRKDPPEFKAKETIQEQFQPSLCHKFI 322
 DB 78 DFRSPSTGLYDNL--EKYHLPPEALFEISYFKKHPEPFALAKELYPGQFPTCHYFM 135
 QY 323 ALSDEKGLLNNTQNTIDLEOVAGIOR--ILOCHGSPATASC--ICKYKVDCEAVRGD 378
 DB 136 RLKDKGGLLCYQNTIDTLERLAGEOEDLVEAGITFTYSHCVASCRHETPLSUMMEK 195
 QY 379 IFNQVPRCPDPADEPLAIKPEIVFGENLPEQFHAMKYDDEVLLIIGSSLVKVR 438
 DB 196 IFSEYTPKCECQ-----SLVKPDIYFGESLIPARFSCMSDPLKVDLLVMGTSLOGR 250
 QY 439 PVA 441
 DB 251 GLA 253
 RESULT 13
 ID 096670 PRELIMINARY; PRT; 351 AA.
 AC 096670;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
 DE SIRT2 HOMOLOG.
 GN SIRT2.
 OS Trypanosoma brucei.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 NCBI_TaxID=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=427;
 RA Hoek M., Wirtz E., Cross G.A.M.;
 RT "Cloning of a Trypanosoma brucei sirt2 homolog."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF102869; AAC73004.1; -
 DR INTERPRO: IPR003000; -
 DR PRAM: PF02146; SIRT2; 1.
 SQ SEQUENCE 351 AA; 38529 MW; 9EAOBD510670575 CRC64;
 Query Match 10.5%; Score 406; DB 5; Length 351;
 Best Local Similarity 30.2%; Pred. No. 7, 8e-21;
 Matches 108; Conservative 57; Mismatches 109; Indels 84; Gaps 11;
 QY 218 NILSEPPKRRKRKRDINTIEDAVKLLD--ECKRIIVLTGAGVSVSGIPDRS--RDGIYAR 274
 DB 10 HVGEP-----TEGLARFERNNTITKIFVAVGASIVAGIPDRSPPTGLYAK 59
 QY 275 LAVDFPDLPDPOAMDIEYFRKDPPEFKFA--KEIYGCQPSLCHKFIASDKEGKIL 332
 DB 60 LS--RYNLSPEDAESLRLRQPSVVFYNIIMDMQIMWGKXCPTTVHNFISLAKKMILL 117

us-09-461-580a-26.rspt

us-09-461-580a-26.rpt

us-09-461-580a-26.rpt

us-09-461-580a-26.rpt

